

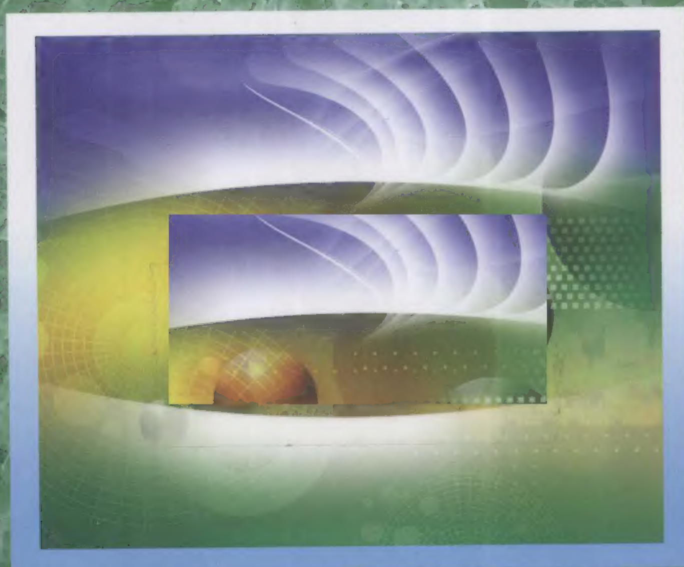
面向“十二五”高等学校精品规划教材

# 数值分析

（第三版）

SHUZHI FENXI

史万明　吴裕树　孙新　编著



北京理工大学出版社
BEIJING INSTITUTE OF TECHNOLOGY PRESS

责任编辑：唐　爽
封面设计：OOICA 原创在线




面向“十二五”高等学校精品规划教材

# 数值分析（第三版）

SHUZHI　FENXI




ISBN 978-7-5640-3107-7

定价：43.00元

面向"十二五"高等学校精品规划教材

# 数值分析

## （第三版）

史万明　吴裕树　孙　新　编著


北京理工大学出版社
BEIJING INSTITUTE OF TECHNOLOGY PRESS

## 内容简介

本书共分10章，内容包括误差知识，方程（组）的迭代解法，线性代数方程组的计算方法，插值法与函数逼近，矩阵的特征值与特征向量的计算方法，数值积分与数值微分，快速傅里叶变换，常微分方程初值问题的数值解法。

全书从构造算法、分析算法、使用算法三方面组织教材内容，力求通俗易懂、深入浅出，并配以例题和习题，以助理解。

本书可作为高等工科院校教材，也可作为工程科技人员的参考书。



图书在版编目（CIP）数据

数值分析/史万明，吴裕树，孙新编著. —3版．—北京：北京理工大学出版社，2010.4

ISBN 978-7-5640-3107-7

Ⅰ.①数… Ⅱ.①史…②吴…③孙… Ⅲ.①数值计算-高等学校-教材 Ⅳ.①O241

中国版本图书馆CIP数据核字（2010）第049819号

出版发行 / 北京理工大学出版社
社　　址 / 北京市海淀区中关村南大街5号
邮　　编 / 100081
电　　话 / (010)68914775(办公室)　68944990(批销中心)　68911084(读者服务部)
网　　址 / http：// www.bitpress.com.cn
经　　销 / 全国各地新华书店
印　　刷 / 天津市建新彩色印刷有限公司
开　　本 / 787毫米×1092毫米　1/16
印　　张 / 21.5
字　　数 / 504千字
版　　次 / 2010年4月第3版　2010年4月第3次印刷
印　　数 / 10001～14000册
定　　价 / 43.00元

责任校对 / 张沁萍
责任印制 / 边心超

# 前　　言

本书是为工科院校的本科生编写的教材，它是在原来教材的基础上，结合多年教学经验和科研实践修订而成的。本着重概念、重方法、重应用、重能力培养的原则，从构造算法、分析算法、使用算法三方面组织教材内容。在构造算法上，除阐明算法的构造思想、原理外，通过进一步地归纳和整理，尽量使同类算法都由某一基本原理或某一基本方法导出，以便读者易于领会和掌握同类算法的共同特征以及同类算法中不同方法之间的相异特征。在分析算法的有关理论推导中，力求深入浅出、通俗易懂，并补充少量基础知识，便于阅读和教学。在算法设计与理论分析中，对每种算法均十分关注其应用条件及使用中的问题。每类算法都配以例题与习题，以助理解和练习。

学习本书所需的数学基础是微积分和线性代数，以及常微分方程的基本概念。读者可针对工科本科生所要求的内容进行选材，其中也包含一部分适合高水平学生深入理解的内容，可供选学。全书共 10 章，约需 70～80 学时，对不同专业，其具体内容和学时数可作适当增减。

本书作者不仅长期从事本门学科的教学，而且具有长期从事科研项目计算的经历，这种实践形成了本书朴素、求实的风格。希望通过本书的介绍，使读者在较短的时间内比较顺利地掌握这些数值方法的要领和基本技巧，为今后从事科学计算打下牢固的基础。

限于水平，书中疏漏和缺陷之处难免，敬请读者批评指正。

编　者

# 目　　录


**第一章　数值计算中的误差** ……（1）

§1　计数与数值 ……（1）

§2　舍入方法与有效数字 ……（7）

§3　算术运算中的误差 ……（10）

§4　算法举例 ……（16）

§5　数值计算中的误差 ……（20）

§6　误差分配原则与处理方法 ……（23）

习题一 ……（27）

**第二章　方程（组）的迭代解法** ……（29）

§1　引言 ……（29）

§2　迭代解法 ……（30）

§3　迭代公式的改进 ……（41）

§4　联立方程组的迭代解法 ……（61）

§5　联立方程组的牛顿解法 ……（68）

§6　联立方程组的延拓解法 ……（70）

习题二 ……（73）

**第三章　解线性方程组的直接法** ……（74）

§1　消元法 ……（74）

§2　选主元的高斯消元法 ……（85）

§3　关于结果精度的检验 ……（87）

习题三 ……（89）

**第四章　解线性方程组的迭代法** ……（90）

§1　向量范数、矩阵范数、谱半径及有关性质 ……（90）

§2　简单迭代法 ……（93）

§3　赛德尔迭代法 ……（99）

§4　松弛迭代法 ……（109）

习题四 ……（115）

**第五章　插值法** ……（117）

§1　不等距节点下的牛顿基本差商公式 ……（117）

§2　等距节点下的牛顿基本差商公式及弗雷瑟图表法 ……（123）

§3 不等距节点下的拉格朗日插值公式 …… (135)
§4 等距节点下的拉格朗日插值公式 …… (138)
§5 插值公式的唯一性及其应用 …… (140)
§6 反插值 …… (142)
§7 埃尔米特插值多项式 …… (150)
§8 三次样条插值 …… (159)
§9 多元函数插值 …… (165)
习题五 …… (169)

**第六章 数值积分和数值微分** …… (172)
§1 数值积分 …… (172)
§2 数值微分 …… (199)
习题六 …… (211)

**第七章 常微分方程数值解法** …… (213)
§1 引言 …… (213)
§2 台劳级数法 …… (214)
§3 基于数值微分公式的方法 …… (215)
§4 龙格-库塔法 …… (216)
§5 线性多步法 …… (221)
§6 单步法的收敛性、相容性与稳定性 …… (234)
§7 差分方程简介 …… (240)
§8 线性多步法的相容性、收敛性与稳定性 …… (242)
§9 方法、阶和步长的选择 …… (246)
§10 常微分方程组和高阶微分方程的数值解法 …… (247)
§11 刚性方程组 …… (251)
§12 对各种方法的比较 …… (253)
习题七 …… (255)

**第八章 函数逼近** …… (256)
§1 离散情况下的最小平方逼近 …… (257)
§2 离散情况下使用正交多项式的最小平方逼近 …… (266)
§3 连续情况下的最小平方逼近 …… (271)
§4 切比雪夫多项式及函数按切比雪夫多项式的展开式 …… (273)
§5 最佳一致逼近 …… (279)
习题八 …… (298)

**第九章 矩阵特征值和特征向量的计算** …… (300)
§1 幂法和反幂法 …… (300)

§2 正交变换矩阵 …………………………………………………………… (307)
§3 雅可比方法 …………………………………………………………… (314)
§4 $QR$ 方法 …………………………………………………………… (319)
习题九…………………………………………………………………… (325)

**第十章 快速傅里叶变换**……………………………………………… (327)
§1 有限离散傅里叶变换 ………………………………………………… (327)
§2 快速傅里叶变换 ……………………………………………………… (329)
习题十…………………………………………………………………… (335)

# 第一章　数值计算中的误差

## §1　计数与数值

### 1.1　远古的计数

数是一串符号或字母的约定性组合，用以表示某种事物的量或值的多寡程度。因此数是事物的量或值的抽象表示，通常称为数值。数值来源于计数，它由远古的计数产生而逐步形成了它的表示方法。计数频繁地在日常生活中出现，无法想象一个成人不会计数。可是人类确实有过一个时期，既不知道用火，也不知道计数。

远古的计数方式现在看不到了，引导我们走向古老年代，帮助我们猜破这个谜的，有以下三条途径。

① 研究语言，研究民间的传说和歌谣。在语言里还保存了许多人类不会写字时代的痕迹。

② 观察婴孩怎样学说话和计数，就像会重演一下人类计数发展的某些步骤，对于人类怎样掌握计数，可以得到一些启示。

③ 研究原始民族。在非洲、南美洲中部以及一些岛屿上，还有一些很落后的部落，与我们五千年前甚至一万年前的祖先差不多，在有些地方还保存着原始生活方式。调查了解后，就能帮助我们知道古时候是怎样计数的。通过以上三个来源的信息，就能大概描绘出我们祖先在发明文字以前是如何计数的。

在人类刚刚学会说话和用火的远古时候，他们只知道两个数：一和二。如果要数的东西不止两个，就简单地说“很多”。近代发现，还有整个部落，数到三就觉得很困难了。在婴孩的发育过程中，也有一段时间，只懂得什么是“一”，什么是“二”，但是不易数到三。慢慢地，又添上了越来越多的新数，人们学会了数到“五”，又把两个“五”加起来成为一个“十”，大自然赋予人类的“计数器”帮助我们学会了它，这个计数器就是两只手上的十个手指。

“五”和“十”这两个数，在计数发展史上起了很大的作用。关于这一点是有许多迹象的。在很多古代民族语言里，前十个数的名称是和手指的名称一样的。在有些现代民族的语言里，也还保存着这个现象的痕迹。例如，在现代意大利语里，“le dita”这个字即表示“到十为止的数字”，也表示“手指”。“屈指一算”也说明早先人类的计数是和手指分不开的。最后，现代的十进制计数法证明了“十”这个数字在计数方法的发展中有多么重大的意义。由此看出，人类首先学会了五个五个地计数，然后把两个五合起来十个十个地计数，中国的算盘就证明了这一点。

在文字出现前，每一件东西，每一个动作都要用一个特别的符号（一个小小的图画）来表示。开始这些图画都较复杂，经过简化形成象形文字，这种象形文字至少用了五千年。那时候还没有特别的符号（数字）来表示数，为了改进计数的技巧，必须在两条路里选择一条：或者是转向用简便的文字，即由象形文字改变到用字母来计数；或者是发明一种方法，采用特别的符号来计数。有的民族走了第一条路，如罗马记数法；另一些民族走了第二条路，如巴比伦记数法和中国记数法。

## 1.2 罗马记数法

字母的发明对于文化的发展有很大的贡献，它也帮助了计数技巧的发展。采用字母来表示数的困难在于，字母是不多的，但是数却有很多。这就是说，不单需要用字母来表示数，而且要发明一种写法，能够用几个字母来写出许多的数。这种用几个字母（或符号）就能写出许多数的方法称作记数法。

罗马记数法特别有意义，因为直到现在，在钟面上、在古老建筑物上都还可以看到它，在书上也还用它来表示章节和世纪等。

古罗马人在几百年中一直使用着一些奇妙的字母来记数，这些字母的起源直到现在还未搞清楚，它们就是

| (1)　V(5)　X (10)　⊥ (50)　)|( (100)

可以设想，表示 1 的字母就是采用一个手指的象形文字，表示 5 的字母就是采用 5 个手指的象形文字



而 10 就是两个 5

VV 或 X → X

但是到了罗马文化最发达的时期（两千年前），这些字母就被和它们相像的拉丁字母代替了。于是改变为

| → I

V → V

X → X

⊥ → L

)|( → C

此外又出现了两个新的字母：D（表示 500）和 M（表示 1 000），其中 C 和 M 可能是拉丁字母 centum(100)和 mille(1 000)的第一个字母。

罗马人又如何写出各个不同的数呢？要写出数字“2”和“3”，他们就简单地把“1”这个字母重复写 2 次和 3 次：Ⅱ，Ⅲ。“4”是这样写的：Ⅳ，在这个写法里，写在 5 左边的一是要从 5 里减去的。反之，写在 5 右边的一是要加到 5 上面的。因此，6、7、8 就写成Ⅵ、Ⅶ、Ⅷ。

再下面就用到Ⅹ这个字母了。“9”写成Ⅸ，接下去是Ⅹ、Ⅺ、Ⅻ、XIII（10、11、12、13)。14 写成 XIV，15 写成 XV 等。20 和 30 就写成几个 10：XX、XXX。要写 40、50 等就要用到字母 L(50)。比如 41 写成 XLI，50、60、70 就写成 L、LX、LXX。要写 90 就使用 C 这个字母，即

XC。100 以内的罗马数字可列写如图 1.1 所示。

| I | II | III | IV | V | VI | VII | VIII | IX | X |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| XI | XII | XIII | XIV | XV | XVI | XVII | XVIII | XIX | XX |
| 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| XXI | XXII | XXIII | XXIV | XXV | XXVI | XXVII | XXVIII | XXIX | XXX |
| 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| XXXI | XXXII | XXXIII | XXXIV | XXXV | XXXVI | XXXVII | XXXVIII | XXXIX | XL |
| 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 |
| XLI | XLII | XLIII | XLIV | XLV | XLVI | XLVII | XLVIII | IL | L |
| 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |
| LI | LII | LIII | LIV | LV | LVI | LVII | LVIII | LIX | LX |
| 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 |
| LXI | LXII | LXIII | LXIV | LXV | LXVI | LXVII | LXVIII | LXIX | LXX |
| 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 |
| LXXI | LXXII | LXXIII | LXXIV | LXXV | LXXVI | LXXVII | LXXVIII | LXXIX | LXXX |
| 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 |
| LXXXI | LXXXII | LXXXIII | LXXXIV | LXXXV | LXXXVI | LXXXVII | LXXXVIII | LXXXIX | XC |
| 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 |
| XCI | XCII | XCIII | XCIV | XCV | XCVI | XCVII | XCVIII | IC | C |
| 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 |

图 1.1

100 后的罗马数字的写法依此类推。102 这个数写成 CII，374 写成 CCCLXXIV 等等。从 400 到 899 间的数都要用到 D(五百)，九百写成 CM，一千写成 M。于是 1 917 这个数就表示为 MCMXVII，1 955 表示为 MCMLV。用罗马数字写出更大的数也不困难，比如 123 849 可写成 CXXIII MDCCCXLIX，其中小写字母“m”表示千，CXXIII m 表示 123 个千。

罗马字母用来记数还可说是方便的，但是用来计算就不方便了，不论哪种算式的演算，要用罗马数字来做，都是几乎不可能的。

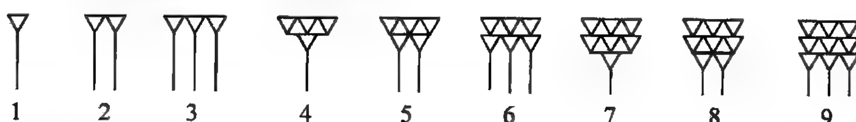
## 1.3　巴比伦记数法

古代巴比伦人用短棒在柔软的黏土土坯上写字，有的烧成了砖，它们形成了“土坯文件”。学者们在发掘古城时找到了许多这样的文件，其中有的是文契、条约、贸易合同或数学文本。用短棒在柔软的土坯上写字，就使巴比伦人所有的象形文字都是由横的或直的楔形▽或∢构成的，因此就叫做“楔形文字”。通过对数以千计的土坯文件的剖析，比较清楚地了解了古代巴比伦人的生活和文化水平。

大约四千年前，在美索不达米亚平原，就是近东的底格里斯河和幼发拉底河流域，即现在的伊拉克国境内，来到了两个游牧民族：苏美尔人和亚克得人，当时他们都是很文明的民族。过了两个世纪，这两个民族就合并成一个强大的国家——巴比伦。在合并前，两个民族都有自己的质量单位和货币单位，苏美尔人的质量单位叫“明那”，货币单位是“明那银子”。亚克得人的单位比较小，其质量单位为“舍克尔”，合明那的六十分之一。在合并后，上述两种质量单位就同时通用了。随着商业与经济的发展，货币流通量也增加了，巴比伦人也需要更大的单位了。很自然，他们又用比明那大 60 倍的单位作为新的质量单位，因为“60”这个数在计数中已很习惯

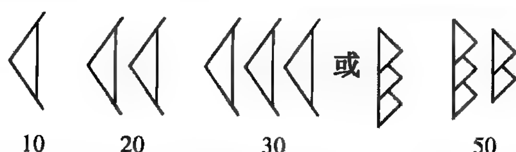
了，这个新单位叫做“塔朗特”；同时也产生了新的货币单位，即一塔朗特银子，它等于 60 明那银子，以上三种单位既是质量单位也是货币单位，每种单位都是较小单位的 60 倍，使得巴比伦人不需要念出和写出比 60 更大的数来了。因此，他们只需要使用 59 个符号来记数。

巴比伦人用直立的楔来表示前 9 个数字




这些符号里的楔排列很合理，念的时候不必去数，因为楔的个数是一眼就可以看出来的。

对于 10、20、30、40、50 的符号是用以下宽度的横写的楔来表示的：




把它们置放在 1～9 个符号的左边就可以获得直至 59 为止的其他记数符号。对于更大的数，就用符号的位置来区分出这些符号是具有什么单位的，它们从右向左的单位分别是舍克尔、明那、塔朗特。比如 2 塔朗特 13 明那 41 个舍克尔就写作

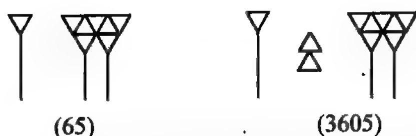



随着社会的发展，又需要记数技术的改进，必须写出越来越大的数，并且这种数用以表示各种各样的量，从而导致数的本身和它所计量的对象脱开而形成了“抽象数”，即不名数，要写出这些抽象数并不需要想出新的符号，只要运用原有的符号就行了。下面写的数



不再像从前那样表示 21 个明那和 32 个舍克尔，而是表示 32 个初级单位(等于 1)加上 21 个第二级单位(等于 60)，这里一个第二级单位有一个初级单位的 60 倍大，更高级的单位依此类推。我们把这种由记号位置不同具有不同单位的记数法叫做进位制记(或计)数法。我们现在还使用着进位制记数法，不同处是我们用的是十进制记数法，而巴比伦人用的是六十进制记数法。

在很长的时期里，巴比伦人没有表示某单位上的符号为零的记号，若出现这种情况，就在该位置上用空位表示。在书写时，这种空位常会因为大小不一致而产生一些混淆。从某一时候起，在巴比伦的楔形文字中出现了一个新的记号——△(分离符)，它相当于零，用来表示一个数里根本不含那一级单位，这样就有




但巴比伦人却没有想到把这个符号放在数字的末尾，因此在巴比伦人的文化里仍然存在混淆不清的数字，如Y表示 1 或 60 或 3 600 等。

虽然巴比伦数学家会写出很大的数，但是还不知数是无限多的。另外，这种记数法尚待进一步完善，在保存其进位制的基础上，采用较小的基数来代替基数“60”以达到简化运算规则的目的。也还要学会准确地使用“零”这个符号。这些是由印度人加以改进和完善的。

## 1.4　印度记数法

15 世纪以后，印度的科学与艺术得到了繁荣，数学特别被尊重，因为可用它来推算历法、确定四季节气的流转，以及预测日、月食等。为了写出很大的数，在印度发明了一种记数法，它把平常习惯的成十的计数和巴比伦人的进位制记数法结合起来，并且巧妙地使用了“零”这个符号。这种记数法后经阿拉伯人带到了欧洲，排除了其他一切记数法而遍及全世界，它就是我们熟悉的十进制记数法，平常叫做“阿拉伯记数法”，正确地说，应该是“印度记数法”。

十进制记数法由低位至高位(由右向左)依次称为个位、十位、百位、千位、万位……在书写和印刷时，每三位叫做一节，节间要留一个小空位(或打上一个“,”)。在数的念法上，世界各民族有所不同，在欧美，对于大数都有专称，英语中称 100 为 hundred，1 000 为 thousand，1 000 000(百万)为 million，从这以后再大的数各国称法有异，美、法称 1 000 000 000(十亿)为 billion，而英、德称1 000 000 000 000(万亿)为 billion。“百万”这个词在欧洲还是近代产生的，是由 13 世纪到中国来的旅行家马可·波罗想出来的，millione 是意大利文字的百万，它是由两个字合成的，其中 mille 是意大利文的“千”，one 是一个字尾，表示“很多”、“很大”。马可·波罗造出这个字来，是为描述“天朝上国”(古时中国的称呼)的无比富庶。

在中国，数字的念法和欧洲各国有所不同，欧洲各国中没有“万”字，“一万”他们叫做“十千”，“十万”叫做“一百千”，到了“一千千”就是百万才有新的名称，上面讲过的数的三位分节就是这个原因。而在中国，千上面还有万、十万、百万、千万，直到“万万”才有一个新名，叫做“亿”。再向上就是十亿、百亿、千亿。到了“万亿”称为“兆”。因此按照中国习惯，数的写法按四位一节划分比较方便。

## 1.5　中国记数法

中国古代，为了计算射猎所获得的飞禽和走兽的数目，就用箭来记录，这就是最初的记数法。后来就逐渐用到其他事物的记数，但箭是很长的，用起来不方便，于是就用短竹子削成筹码来代替箭，从这里就产生了象形文字的数字。中国的数字分成横、竖两种形式



注意：到 5 以上，就用一根筹码来代替 5，因为筹码用得太多就看不清了。为何到了 5 才用另一根筹码来代替呢？这当然与我们手上有 5 个手指分不开的。到了 10 以上，就不再用新的记号而采用进位制，例如 12 就写成|=或—‖。所以中国古代的记数法，可以说是“用五小进，用十大进”，直到近代也还使用着“一五一十”的口诀。大约到了 6 世纪，人们开始用珠来代替筹码，出现了原始的算盘，到了元代，算盘已在全国风行。从这里看来，中国记数法和巴比伦记数法一样，是采用象形文字和进位制的。书写时采取横竖相间的方式以免混淆不清，例如 1 289 就写成|=Ⅲ≜。开始中国也没有零，遇到零就空一位。到后来，为简化书写，又把其中几个复杂的数字改变为

≣ → X，≣ → ō 或 ȯ，≣ → X̄ 或 Ẋ

以后ȯ⟶8，X̄或Ẋ⟶Ẋ，而竖式的六、七、八已不再采用，并且造出了 0 这个记号，最后成为近代习惯用的数字

| 一<br>\| | 二<br>‖ | 三<br>Ⅲ | X | 8 | ⊥ | ⊥̲ | ≜ | Ẋ | O |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |

中国数字的好处，在于容易使人一看便知道各数字所代表的数，另外它也采用十进制。缺点是书写起来不太方便，有的数字要写好几笔。

## 1.6 通用的记数法

通用的记数法是以印度记数法为原理的 $R$ 进制记数法。其定点形式（称为定点数）为

$$
\begin{aligned}
&(\alpha_n\alpha_{n-1}\cdots\alpha_1\alpha_0.\beta_1\beta_2\cdots\beta_m)_R \\
&= \alpha_nR^n + \alpha_{n-1}R^{n-1} + \cdots + \alpha_1R^1 + \alpha_0R^0 + \frac{\beta_1}{R^1} + \frac{\beta_2}{R^2} + \cdots + \frac{\beta_m}{R^m}
\end{aligned}
\tag{1.1}
$$

式中，每位数字 $\alpha_i,\beta_j(i=0,1,2,\cdots,n;j=1,2,\cdots,m)$ 是介于 0 与 $R-1$ 间的正整数。式(1.1)左边是数的书写形式，右边是该数所表达的值，它等于每位数字与其单位乘积之和。该数的总位数称为字长。

通用记数法的浮点形式（称为浮点数）为

$$
R^p\cdot(0.\ d_1\,d_2\cdots d_m)_R = R^p\cdot\left(\frac{d_1}{R^1}+\frac{d_2}{R^2}+\cdots+\frac{d_m}{R^m}\right) \tag{1.2}
$$

式中，$p$ 为阶码，$(0.\ d_1\,d_2\cdots d_m)_R$ 称为尾数，这里 $d_i$ 为介于 0 与 $R-1$ 间的正整数。若 $d_1\neq 0$，则称该浮点数为规格化数；否则称为非规格化数。尾数的位数 $m$ 称为字长。在下面三数中

$$
(37.218\ 29)_{10} = 10^2\cdot(0.372\ 182\ 9)_{10} = 10^3\cdot(0.037\ 218\ 29)_{10}
$$

左边是定点数，中间是浮点规格化数，右边是浮点非规格化数。

数的书写形式与其表达的值经常不加区分地统称为数值。采用浮点形式表达数，可以拓宽数值的表达范围，特别是对于甚大（如天文数字）或甚小（如侏儒数字）的定点数可化简其写法。例如，地球的质量是 6 000 000 000 000 000 000 000t 可表为 $10^{22}\cdot(0.6)_{10}$t。氢原子的质

量是0.000 000 000 000 000 000 000 001 65g 可表为$10^{-23}\cdot(0.165)_{10}$g。

进位制的基数$R$愈大，则数的字长愈短，但其运算规则却随着每位中数字符号的增加而变得更加复杂；反之亦然。

伴随数值的产生和社会发展的需要，又产生了对数值的各种运算及实现运算的种种计算工具，在运算的基础上又产生了各种数值方法，运用它们去解决科学研究或工程技术中的实际问题。特别是电子数字计算机的诞生，是计算数学史甚至人类文明史上的一个里程碑，它使人类获得了高速度、自动化的计算工具，为众多浩繁的数值计算问题的解决展现了光明的前景。同时，也自然地促进数值方法的迅速发展和更新。目前电子数字计算机已成为广泛应用的数值计算工具，但本质上它们能执行的也只是算术运算（加减乘除四则运算）和逻辑运算。而数学问题中的运算范围则是极为广阔的，除算术运算外，还有代数运算、函数运算以及其他种种复杂的运算。数值分析研究的问题，就是为求解各类数学问题去构造算法、分析算法和使用算法。所谓算法就是由基本运算及规定的运算顺序所构成的完整的解题步骤。构造算法，应以计算机所能执行的运算为依据，尽可能节省存贮量、计算量和提高计算精度。分析算法，就是在数值计算类问题中，主要分析算法的收敛性、稳定性和误差估计等；在逻辑计算类问题中，主要研究算法的时间复杂性和空间复杂性，这部分内容已有另一门学科“算法复杂性”专门讲述，它与前一类问题是相互关联的，限于篇幅，本书不多涉及这部分内容。使用算法，就是要注意算法的应用条件、计算过程的控制以及计算机上的使用问题等。

# §2　舍入方法与有效数字

数值除来源于计数外，还大量地来源于测量。由于测量工具本身具有的精确度不同，所得到的测量值只能近似地反映出所测物理量的大小。为了衡量其近似的程度，我们引入以下两类误差：绝对误差和相对误差。

## 2.1　绝对误差与相对误差

设$A$为精确值，$a$为近似值，则定义它们之差

$$\Delta = a - A \tag{1.3}$$

为近似值$a$的绝对误差，简称误差[1]。当$\Delta>0$时，称为正绝对误差；否则称为负绝对误差。由于精确值一般是未知的，因而$\Delta$不能求出来。但根据测量误差或计算的情况可以估计出它的上界$\varepsilon$

$$|\Delta| = |a - A| < \varepsilon \tag{1.4}$$

则称$\varepsilon$为$a$的绝对误差限或误差限。由上式可以推知精确值所界定的范围为

$$a - \varepsilon < A < a + \varepsilon \tag{1.5}$$

有时也用

$$A = a \pm \varepsilon \tag{1.6}$$

来表示。

要刻画近似值的精确程度还必须考虑该精确值本身的大小，以便比较单位数值中所含有

[1] 误差亦可定义为$\Delta=A-a$，这时误差的符号与本定义相反。

的误差，这就导致引入相对误差的概念。相对误差定义为绝对误差与精确值之比

$$\delta = \Delta/A \tag{1.7}$$

因 $A$ 一般不知道，实际计算时采用下式

$$\delta = \Delta/a \tag{1.8}$$

来代替，这样代替后，其误差为

$$\frac{\Delta}{A} - \frac{\Delta}{a} = \frac{a-A}{Aa}\Delta = \frac{1}{Aa}\Delta^2$$

当 $\Delta$ 很小时，上述误差为 $\Delta$ 的高阶无穷小，因此式(1.8)的取法是合理的。相对误差绝对值之上界 $\eta$

$$|\Delta/a| < \eta \tag{1.9}$$

称为相对误差限。例如

① $A=0.3\times10^1$，$a=0.31\times10^1$，则 $\Delta=0.1$，$\delta=0.\dot{3}\times10^{-1}$

② $A=0.3\times10^{-3}$，$a=0.31\times10^{-3}$，则 $\Delta=0.1\times10^{-4}$，$\delta=0.\dot{3}\times10^{-1}$

③ $A=0.3\times10^4$，$a=0.31\times10^4$，则 $\Delta=0.1\times10^3$，$\delta=0.\dot{3}\times10^{-1}$

上例表明，计算出的绝对误差虽然差别很大，但它们却有相同的相对误差。显然，对精确性的衡量，只讨论绝对误差是不够的。

绝对误差是有量纲的量，而相对误差是一个无量纲的量，有时亦用百分比、千分比等来表示。一般量值范围小的场合以采用绝对误差限为多；量值范围大的场合则采用相对误差限较多。亦有兼有两者的情况，例如在炮兵作业中，采用以下准则来控制对目标的射击命中率：

当 $d_{pm}\leqslant 20\,000$ m，要求 $|d_{zm}|\leqslant 10$ m

当 $d_{pm}>20\,000$ m，要求 $|d_{zm}|/|d_{pm}|\leqslant 5/10\,000$

式中，$d_{pm}$ 为炮阵地与目标间的水平距离；$d_{zm}$ 为炸点与目标间的水平距离。

## 2.2 舍入方法

由于我们只能取数值的有限位字长进行计算，因此在计算前，对于一个无限位字长的精确数或字长较长的近似数必须处理成有限位字长的近似数，这种处理方法称为舍入方法。设待舍入处理的数为

$$A = a_0a_1\cdots a_m.\,a_{m+1}a_{m+2}\cdots a_{m+n}a_{m+n+1}\cdots \qquad (a_0\neq 0) \tag{1.10}$$

今要求对 $A$ 作舍入处理，以获得具有 $n$ 位小数的近似数 $a$，下面讨论不同的舍入方法。

### 2.2.1 截断法

我们在 $A$ 中的数字 $a_{m+n}$ 与 $a_{m+n+1}$ 间将 $A$ 切分为高位部分与低位部分

$$A = a_0a_1\cdots a_m.\,a_{m+1}\cdots a_{m+n}(\text{高位部分}) + 0.\underbrace{0\cdots0}_{n\text{位}}a_{m+n+1}\cdots(\text{低位部分}) \tag{1.11}$$

截断法就是截取 $A$ 的高位部分作为近似数 $a$

$$a = a_0a_1\cdots a_m.\,a_{m+1}\cdots a_{m+n} \tag{1.12}$$

其舍入误差限估计如下

$$|\Delta| = |a-A| = 0.\overbrace{0\cdots0}^{n\text{位}}a_{m+n+1}\cdots$$

$$
\begin{aligned}
&\leqslant 0.\overbrace{0\cdots 0}^{n\text{位}}\dot{9} \\
&\leqslant 0.0\cdots 1 = 1\times 10^{-n}
\end{aligned} \tag{1.13}
$$

由上式可见，这种舍入方法导致的舍入误差限不超过近似数 $a$ 最末位数字的一个单位，具有这种精度的近似数 $a$ 称为准确到小数后第 $n$ 位的可靠数，其每一位数字均称为可靠数字。

### 2.2.2　四舍五入法

此法根据低位部分的最高位数字 $a_{m+n+1}$ 的大小对高位部分的最末位数字 $a_{m+n}$ 进行适当的修改，使 $a$ 的绝对误差限具有最小值，具体方法如下。

(1) 四舍情况

当 $a_{m+n+1}=1,2,3,4$ 时，取近似值 $a$ 为

$$
a = a_0 a_1 \cdots a_m . a_{m+1} \cdots a_{m+n} \tag{1.14}
$$

其舍入误差限为

$$
\begin{aligned}
|\Delta| = |a-A| &= 0.\overbrace{0\cdots 0}^{n\text{位}} a_{m+n+1}\cdots \\
&\leqslant 0.0\cdots 4\dot{9} \\
&\leqslant 0.0\cdots 5 = 0.5\times 10^{-n}
\end{aligned} \tag{1.15}
$$

(2) 五入情况

当 $a_{m+n+1}=5,6,7,8,9$ 时，取近似值 $a$ 为

$$
a = a_0 a_1 \cdots a_m . a_{m+1} \cdots (a_{m+n}+1) \tag{1.16}
$$

其舍入误差限为

$$
\begin{aligned}
|\Delta| = |a-A| &= |\,0.\overbrace{0\cdots 1}^{n\text{位}} - 0.\overbrace{0\cdots 0}^{n\text{位}} a_{m+n+1}\cdots\,| \\
&\leqslant |0.\ 0\cdots 1 - 0.0\cdots 05| = 0.5\times 10^{-n}
\end{aligned} \tag{1.17}
$$

综合(1)、(2)可知，用四舍五入法所得的近似值 $a$，其舍入误差限不超过 $0.5\times10^{-n}$，即其最末位数字的半个单位，与截断法比较，其舍入误差限缩小了一半。

### 2.2.3　改进的四舍五入法

在四舍五入法中，显然数字 $a_{m+n+1}$ 在五入情况下的个数较四舍情况下的个数多 1，而在大量运算中，数字 1～9 在 $a_{m+n+1}$ 位上出现的次数大体上是相同的。因此按四舍五入法对大量运算结果作舍入处理，有可能导致最终结果的数值偏大的弊病。为改进它，可对四舍五入法附加以下补充规定，方法如下。

奇进偶不进法：

$$
a_{m+n+1}\begin{cases} <5，将\ a_{m+n}\ 后的数字舍去 \\ >5，对\ a_{m+n}\ 作加\ 1\ 处理 \\ =5，\begin{cases} 当\ a_{m+n}\ 为奇数时，对\ a_{m+n}\ 作加\ 1\ 处理 \\ 当\ a_{m+n}\ 为偶数时，将\ a_{m+n}\ 后的数字舍去 \end{cases} \end{cases} \tag{1.18}
$$

偶进奇不进法：

$$
a_{m+n+1}\begin{cases} <5，将\ a_{m+n}\ 后的数字舍去 \\ >5，对\ a_{m+n}\ 作加\ 1\ 处理 \\ =5，\begin{cases} 当\ a_{m+n}\ 为奇数时，将\ a_{m+n}\ 后的数字舍去 \\ 当\ a_{m+n}\ 为偶数时，对\ a_{m+n}\ 作加\ 1\ 处理 \end{cases} \end{cases} \tag{1.19}
$$

以上两种方法都能达到进、舍的几率相同，但采用奇进偶不进法更为有利，这是由于作这种舍入处理后的数值均是偶数，一般而论，能把偶数恰好除尽的数要比能把奇数恰好除尽的数要多，这有助于提高计算结果的精度。实践证明，在大量运算中，按上述改进的方法作舍入处理，整个运算过程的舍入误差积累较小。

### 2.3 有效数字

前面已经证明，通过四舍五入法得到的近似值，其舍入误差限不超过 $0.5\times10^{-n}$，即其最末位数字的半个单位，具有这种精度的近似数 $a$ 称为准确到小数后第 $n$ 位的有效数，其每一位数字均称为有效数字。如果一个数是有效的，则可立即获得该数关于其绝对误差限和相对误差限的估计如下。

**结论 1** 对于给出的一个有效数，其绝对误差限不大于其最末位数字的半个单位。

**结论 2** 对于给出的一个有效数，其相对误差限可估计如下：

$$|\delta|=\left|\frac{0.5\times10^{-n}}{a_0\times10^{m}+a_1\times10^{m-1}+\cdots}\right| \leqslant\left|\frac{0.5\times10^{-n}}{a_0\times10^{m}}\right|=\frac{5}{a_0}\times10^{-(m+n+1)} \tag{1.20}$$

因近似数 $a$ 具有$(m+n+1)$位有效数字，由式(1.20)可见，有效数位愈多，其相对误差就愈小。显见，要想缩小相对误差，最直接而有效的办法就是增加运算中的有效位数。

注意：

① 近似数 $a$ 的有效数字应从其左边第一个不等于零的数字开始计数，直至最末位数字为止。该数前面的零不计为有效数字，但后面的零应为有效数字。如 0.001 000 为经四舍五入后的有效数，其有效数位为 4 位。

② 浮点数的有效数应由其定点部分的有效数位确定。如 $a=75\times10^{-3}$，其定点部分 75 具有 2 位有效数字，所以 $a$ 为具有 2 位有效数字的浮点数，其绝对误差限为 $|\Delta a|=(0.5\times10^{0})\times10^{-3}=0.000\,5$，相对误差限为 $\delta a=0.000\,5/75\times10^{-3}=0.000\,\dot{6}$。

③ 若已知数 $a$ 及其绝对误差限 $|\Delta a|$，要求对 $a$ 作舍入处理并确定其有效数位。这时可将 $|\Delta a|$ 扩大成 $|\Delta a|\leqslant0.5\times10^{-k}$ 的形式，然后对 $a$ 作截止到小数 $k$ 位的舍入处理。如近似数 $a=1.234\,5$，已知其绝对误差限 $|\Delta a|\leqslant0.000\,83$，因为 $0.000\,83<0.005=0.5\times10^{-2}$，所以可将 $a$ 舍入处理为 $a=1.23$，它具有 3 位有效数字。

④ 若要求近似数 $a$ 的绝对误差限小于 $\varepsilon$，问 $a$ 应取几位有效数字？这时可据 $\varepsilon$ 写出以下不等式 $0.5\times10^{-k}\leqslant\varepsilon$，由此确定出 $a$ 应取至小数后 $k$ 位。如近似数 $a$ 的绝对误差限要求小于 $\varepsilon=0.003$，问 $a$ 应取几位有效数字？因为 $0.000\,5<0.003$，所以可取 $|\Delta a|=0.000\,5=0.5\times10^{-3}$，可见 $a$ 的有效数位至少应取至小数后 3 位为止。

## §3 算术运算中的误差

所谓算术运算指的是加、减、乘、除这四种基本运算，其他的运算都可通过一定的算法化为一系列算术运算来完成。这里，我们来考虑数据误差在算术运算中的传播规律并对结果的误差进行估计。

设 $x^*$、$y^*$ 为准确值，$x$、$y$ 分别为其近似值。则它们的绝对误差分别为 $\Delta x=x-x^*$，$\Delta y=y-y^*$。对于 $\Delta x$ 和 $\Delta y$ 常用其主部(指数值的高位部分)近似它们：$\mathrm{d}x\approx\Delta x$，$\mathrm{d}y\approx\Delta y$，它们之间的差别只体现在数值的低位部分，其值甚微可以略去。因此，对于近似值间的算术运算所产生的结果误差的主部可按微分公式来近似估算。

### 3.1　$c=x\pm y$

$$|\mathrm{d}c|=|\mathrm{d}x\pm\mathrm{d}y|\leqslant|\mathrm{d}x|+|\mathrm{d}y|\leqslant\varepsilon_x+\varepsilon_y \tag{1.21}$$

其中
$$|\Delta x|=|x-x^*|\leqslant\varepsilon_x,\ |\Delta y|=|y-y^*|\leqslant\varepsilon_y$$

**例 1.1**　求近似值 285.35，196.87，58.43，4.96 的和，其中每个数的绝对误差限为 $0.5\times10^{-2}$。

**解**　$285.35+196.87+58.43+4.96=545.61$

和 545.61 的绝对误差限为

$$4\times(0.5\times10^{-2})=0.02<0.5\times10^{-1}$$

因此和值 545.61 应去伪存真作舍入处理成 545.6，它具有 4 位有效数字。

**例 1.2**　求有效数 3.150 950，15.426 463，568.375 8，7 684.388 的和。

**解**　$3.150\,950+15.426\,463+568.375\,8+7\,684.388=8\,271.341\,213$

和 8 271.341 213 的绝对误差限为

$$2\times(0.5\times10^{-6})+0.5\times10^{-4}+0.5\times10^{-3}\approx0.5\times10^{-3}$$

因此和值 8 271.341 213 应舍入成 8 271.341。由此可见，和值 8 271.341 213 中最末 3 位的计算工作是没有意义的。合理的做法是将小数位数较多的各数按小数位数最少的位数多取 1 位作舍入处理，在本例中是舍成四位小数后再相加

$$3.151\,0+15.426\,5+568.375\,8+7\,684.388=8\,271.341\,3$$

则和 8 271.341 3 的绝对误差限为

$$3\times(0.5\times10^{-4})+0.5\times10^{-3}=0.000\,65<0.5\times10^{-2}$$

和值 8 271.341 3 舍入至小数后二位得 8 271.34。

注意：

① 大量运算时，$\sum\varepsilon_i$ 可能很大。

② 两个相差很大的数进行加减时，要防止大数“吃”小数现象。在电子数字计算机上，浮点数的值有一定的数值范围。如果数值小于该范围内的最小数，机器将它置为 0；反之，若数值大于该范围内的最大数，机器将视为无穷大而中断计算。当两个浮点数进行加减时，要对齐小数点(称为对阶)，它是将小阶向大阶看齐而将具有小阶数的尾数作相应右移来实现的。对阶的结果，可能会把阶码小的数的尾数部分或全部移掉，称大数“吃”小数现象。例如，在四位浮点计算机上运算下述加法：

$10^3(0.896\,1)+10^{-3}(0.468\,8)$

$\longrightarrow 10^3(0.896\,1)+10^3(0.000\,0)004\,688$ (对阶：小数阶码加 6，尾数右移 6 位)

$\longrightarrow 10^3(0.896\,1)$

其结果大数吃掉了小数。

基于上述原因，$(a+b)+c$ 可能不等于 $(a+c)+b$，例如 $a=10^{12}$，$b=10^1$，$c=-a$ 时，在八位

$A'Ae_i=\lambda_i e_i$，则对任何$\|X\|_2=1$的$X$可表为

$$X = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n$$

计算

$$\begin{aligned}\|AX\|_2^2 &= (AX)'(AX) = X'(A'A)X \\ &= (x_1 e_1 + \cdots + x_n e_n)'(A'A)(x_1 e_1 + \cdots + x_n e_n) \\ &= (x_1 e_1 + \cdots + x_n e_n)'(\lambda_1 x_1 e_1 + \cdots + \lambda_n x_n e_n) \\ &= (x_1 e'_1 + \cdots + x_n e'_n)(\lambda_1 x_1 e_1 + \cdots + \lambda_n x_n e_n) \\ &= \lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2 \\ &\leqslant \lambda_1 (x_1^2 + x_2^2 + \cdots + x_n^2) \qquad (\lambda_1 = \max_i \lambda_i) \\ &= \lambda_1\end{aligned}$$

另一方面，取$X=e_1$，则$\|AX\|_2^2=\lambda_1$，即上界可达。从而证得

$$\|A\|_2^2 = \max_{\|X\|=1} \|AX\|_2 = \lambda_1 = \lambda_{\max}(A'A) = \rho(A'A)$$

特别当$A$为对称矩阵时，有$A'=A,A'A=A^2$，记$A$的特征值为$\mu_1,\mu_2,\cdots,\mu_n$，且$|\mu_1|\geqslant|\mu_2|\geqslant\cdots\geqslant|\mu_n|$，则有$\lambda_i=\mu_i^2$（因为$A'A=A^2$），就有

$$\|A\|_2 = \sqrt{\lambda_{\max}(A'A)} = \sqrt{\lambda_{\max}(A^2)} = \sqrt{\max_{1\leqslant i\leqslant n}\mu_i^2} = \max_{1\leqslant i\leqslant n} |\mu_i| = \rho(A) = |\mu_1|$$

（证毕）

由于2—范数具有关系式(4.20)，所以$\|A\|_2$被称为谱范数。

**定理4.5**　设$A$是任意$n$阶方阵，由$A$的各次幂所组成的矩阵序列

$$I,\ A,\ A^2,\ \cdots,\ A^k,\cdots \tag{4.21}$$

收敛于零，则$\lim\limits_{k\to\infty}A^k=0$的必要充分条件是

$$\rho(A) < 1 \tag{4.22}$$

证明略。

本章叙述最常用的几种迭代法，包括简单迭代法、赛德尔迭代法、松弛迭代法以及迭代法的收敛性与精度控制的问题。使用迭代法求解线性方程组，具有计算简单、编制程序容易、存储量较小、舍入误差积累小(只需计算最终迭代那一次的舍入误差)等优点，较适合于高阶线性方程组的求解。

# §2　简单迭代法

## 2.1　迭代公式

对于线性方程组$AX=B$，首先应将其改写为$X=\Phi(X)$的形式，其方法是多种多样的。下面介绍两种常用的迭代格式。

### 2.1.1　迭代格式1

将$AX=B$改写为$0=B-AX$，两边加上$X$后得

$$X = [I-A]X + B = CX + B \tag{4.23}$$

或写成

$$x_i = \sum_{j=1}^{n} c_{ij} x_j + b_i \qquad (i = 1,2,\cdots,n) \tag{4.24}$$

若$\varepsilon_x=\varepsilon_y=\varepsilon$，则上式成为

$$|\mathrm{d}c|\leqslant(|x|+|y|)\varepsilon \tag{1.27}$$

相对误差限为

$$\begin{aligned}|\delta c|&=\left|\frac{\mathrm{d}c}{c}\right|=\left|\frac{x\mathrm{d}y+y\mathrm{d}x}{xy}\right|\\&=\left|\frac{\mathrm{d}x}{x}+\frac{\mathrm{d}y}{y}\right|=|\delta x+\delta y|\leqslant|\delta x|+|\delta y|\end{aligned} \tag{1.28}$$

上式说明，乘积的相对误差等于各乘数之相对误差之和，其相对误差限等于各乘数相对误差限之和。

**例 1.3**　求$c=12.2\times73.56$的绝对误差限和相对误差限。

**解**　$c=12.2\times73.56=897.432$

$$|\delta c|\leqslant\frac{0.5\times10^{-1}}{12.2}+\frac{0.5\times10^{-2}}{73.56}=0.004\,2$$

$$|\mathrm{d}c|\leqslant897.432\times0.004\,2=3.77<5=0.5\times10^{1}$$

按$|\mathrm{d}c|$取$c$为$90\times10^{1}$。

### 3.3　$c=\dfrac{x}{y}$

$$|\mathrm{d}c|=\left|\frac{y\mathrm{d}x-x\mathrm{d}y}{y^2}\right| \tag{1.29}$$

$$\begin{aligned}|\delta c|&=\left|\frac{y\mathrm{d}x-x\mathrm{d}y}{y^2\cdot\dfrac{x}{y}}\right|=\left|\frac{\mathrm{d}x}{x}-\frac{\mathrm{d}y}{y}\right|\\&\leqslant\left|\frac{\mathrm{d}x}{x}\right|+\left|\frac{\mathrm{d}y}{y}\right|=|\delta x|+|\delta y|\end{aligned} \tag{1.30}$$

上式说明商的相对误差限等于除数与被除数的相对误差限之和。

**例 1.4**　求$c=\dfrac{25.7}{3.6}$的绝对误差限与相对误差限。

**解**　$c=\dfrac{25.7}{3.6}=7.138\,89$

$$|\delta c|\leqslant\frac{0.5\times10^{-1}}{25.7}+\frac{0.5\times10^{-1}}{3.6}=0.016$$

$$|\mathrm{d}c|\leqslant7.138\,89\times0.016=0.11<0.5\times10^{0}$$

按$|\mathrm{d}c|$取$c=7$。

注意：

① 当分母$y^*$很小时，$|\mathrm{d}c|$可能很大。

$$c=\frac{x}{y^*}=\frac{x^*+\Delta x}{y^*}=\frac{x^*}{y^*}+\frac{\Delta x}{y^*}=\text{商的真值}+\mathrm{d}c$$

$$\mathrm{d}c=\frac{1}{y^*}\cdot\Delta x$$

当$y^*$很小时，$1/y^*$就很大，它对分子的误差$\Delta x$有极大的放大作用。因此要尽量避免被除数绝对值远远大于除数绝对值的除法和绝对值很大的乘数的乘法。

② 当分母为两个相近数相减时，常会因有效数位的丧失而出现①的情况。例如

$$\frac{\text{分子}}{0.1456-0.1455}=\frac{\text{分子}}{0.0001}=10^4\cdot(\text{分子})$$

中，分子的误差可被扩大 $10^4$ 倍。

### 3.4 $c=x^p(p>1)$

$$|\mathrm{d}c| = |px^{p-1}\mathrm{d}x| \tag{1.31}$$

$$|\delta c| = \left|\frac{px^{p-1}\mathrm{d}x}{x^p}\right| = p\left|\frac{\mathrm{d}x}{x}\right| = p|\delta x| \tag{1.32}$$

由上式可见，$x$ 的 $p$ 次幂的相对误差是 $x$ 本身的相对误差的 $p$ 倍。

若令 $p=\dfrac{1}{q}$，则得

$$c=\sqrt[q]{x}$$

$$|\delta c| = \frac{1}{q}\left|\frac{\mathrm{d}x}{x}\right| = \frac{1}{q}|\delta x| \tag{1.33}$$

即 $x$ 的 $q$ 次根的相对误差是 $x$ 本身的相对误差的 $1/q$ 倍。

**例 1.5** 求 $c=(12.2)^2$ 的绝对误差限和相对误差限。

**解**

$$c=(12.2)^2=148.84$$

$$|\delta c|\leqslant 2\times\frac{0.5\times10^{-1}}{12.2}=0.0082$$

$$|\mathrm{d}c|\leqslant 148.84\times0.0082=2.44<5=0.5\times10^1$$

按 $|\mathrm{d}c|$ 取 $c=15\times10^1$。

**例 1.6** 设正方形面积为 $s=12.34$，其绝对误差限为 $|\Delta s|\leqslant0.01$，问边长 $a$ 具有多大的相对误差和多少位有效数字？

**解** 因

$$a=\sqrt{s}=\sqrt{12.34}=3.5128\cdots$$

$$|\delta s|=\frac{0.01}{12.34}=0.0008$$

则

$$|\delta a|\leqslant\frac{1}{2}|\delta s|=0.0004$$

要求

$$|\mathrm{d}a|=3.5128\times0.0004=0.0014$$

因为 $0.0005<0.0014$

所以可取 $|\mathrm{d}a|=0.0005=0.5\times10^{-3}$

按 $|\mathrm{d}a|$ 应取 $a=3.513$，它具有 4 位有效数字。

### 3.5 数学问题解的误差估计

对于不同的数学问题，它们的解与参量（原始数据）$x_1,x_2,\cdots,x_n$ 有关，我们把数学问题的函数关系形式地记为

$$y = f(x_1,x_2\cdots,x_n) \tag{1.34}$$

这种对应关系有的可用解析式表出，有的则以方程的形式隐含地给出。当参量有误差 $\Delta x_i(i=1,2,\cdots,n)$ 时，必定引起解的误差 $\Delta f$，假定 $f$ 在点 $(x_1,x_2\cdots,x_n)$ 可微，则当 $\Delta x_i(i=1,2,\cdots,n)$ 较小时，解的误差限可估计如下

$$
\begin{aligned}
|\mathrm{d}f(x_1,x_2,\cdots,x_n)| &\approx \left|\sum_{i=1}^{n}\frac{\partial f(x_1,x_2,\cdots,x_n)}{\partial x_i}\mathrm{d}x_i\right| \\
&\leqslant \sum_{i=1}^{n}\left|\frac{\partial f(x_1,x_2,\cdots,x_n)}{\partial x_i}\right|\cdot|\mathrm{d}x_i| \\
&\leqslant \sum_{i=1}^{n}\left|\frac{\partial f(x_1,x_2,\cdots,x_n)}{\partial x_i}\right|\cdot\varepsilon_i
\end{aligned}
\tag{1.35}
$$

其中$|\mathrm{d}x_i|\leqslant\varepsilon_i,(i=1,2,\cdots,n)$。解的相对误差限为

$$
\begin{aligned}
|\delta f(x_1,x_2,\cdots,x_n)| &\approx \left|\sum_{i=1}^{n}\frac{\partial f(x_1,x_2,\cdots,x_n)}{\partial x_i}\cdot\frac{x_i}{f(x_1,x_2,\cdots,x_n)}\cdot\frac{\mathrm{d}x_i}{x_i}\right| \\
&\leqslant \sum_{i=1}^{n}\left|\frac{\partial f(x_1,x_2,\cdots,x_n)}{\partial x_i}\cdot\frac{x_i}{f(x_1,x_2,\cdots,x_n)}\right|\cdot\delta_i
\end{aligned}
\tag{1.36}
$$

其中

$$
\left|\frac{\mathrm{d}x_i}{x_i}\right|\leqslant\delta_i \qquad (i=1,2,\cdots,n)
$$

$$
\begin{cases}
A_i = \left|\dfrac{\partial f(x_1,x_2,\cdots,x_n)}{\partial x_i}\right| \\
B_i = \left|\dfrac{\partial f(x_1,x_2,\cdots,x_n)}{\partial x_i}\cdot\dfrac{x_i}{f(x_1,x_2,\cdots,x_n)}\right|
\end{cases}
\tag{1.37}
$$

$A_i$ 和 $B_i$ 分别表示对参量的绝对误差和相对误差的放大或缩小的倍数，它们的大小可以用来衡量解对参量误差的敏感程度。

需要指出，式(1.35)、式(1.36)仅当 $\mathrm{d}x_i(i=1,2,\cdots,n)$较小时才是可用的。若较大，则 $\mathrm{d}f$ 或 $\delta f$ 按 $\mathrm{d}x_i$ 进行线性叠加的估计与实际上为非线性变化间的差别会增大，即 $\mathrm{d}f$ 或 $\delta f$ 按各 $\mathrm{d}x_i$ 为独立性影响累计与实际上是关联性影响的综合累计间的差别会增大。在这种情况下，按上述微分法的误差估计就十分不可靠了。

**例 1.7**　已知球体的直径 $D=3.7$ cm，按

$$
V=\frac{1}{6}\pi D^3 \tag{1.38}
$$

计算其体积，求其绝对误差限与相对误差限。

**解**　若取 $\pi=3.14$，计算式(1.38)得

$$
V=\frac{1}{6}\times 3.14\times 3.7^3=26.5 \tag{1.39}
$$

按式(1.35)得

$$
|\mathrm{d}V|\leqslant\left|\frac{\partial V}{\partial \pi}\right|\cdot|\mathrm{d}\pi|+\left|\frac{\partial V}{\partial D}\right|\cdot|\mathrm{d}D| \tag{1.40}
$$

式中

$$
\frac{\partial V}{\partial \pi}=\frac{1}{6}D^3=\frac{1}{6}\times 3.7^3=8.44
$$

$$
\frac{\partial V}{\partial D}=\frac{1}{2}\pi D^2=\frac{1}{2}\times 3.14\times 3.7^2=21.5
$$

$$
|\mathrm{d}\pi|\leqslant 0.001\,6,\ |\mathrm{d}D|\leqslant 0.5\times 10^{-1}
$$

代入式(1.40)得

$$
|\mathrm{d}V|\leqslant 8.44\times 0.001\,6+21.5\times(0.5\times 10^{-1})=1.088\approx 1.1
$$

$$\left|\frac{\mathrm{d}V}{V}\right| \leqslant \frac{1.088}{26.5} = 0.04 = 4\%$$

## §4 算法举例

**例 1.8** 计算

$$D = \frac{0.000\,5 \times 0.014\,3 \times 0.001\,2}{0.000\,3 \times 0.012\,5 \times 0.013\,5} \tag{1.41}$$

**解** 算法 1 分子分母分别计算后相除(各取 9 位小数)。

$$\begin{aligned} A &= 0.000\,5 \times 0.014\,3 \times 0.001\,2 \\ &= 0.000\,007\,15 \times 0.001\,2 \\ &= 0.000\,000\,009 (\text{有舍入}) \\ B &= 0.000\,3 \times 0.012\,5 \times 0.013\,5 \\ &= 0.000\,003\,75 \times 0.013\,5 \\ &= 0.000\,000\,051 (\text{有舍入}) \end{aligned}$$

$$D_1 = \frac{A}{B} = 0.176\,47$$

$$|\delta A| \leqslant \frac{0.000\,000\,000\,5}{0.000\,000\,009} = 0.055\,6$$

$$|\delta B| \leqslant \frac{0.000\,000\,000\,5}{0.000\,000\,051} = 0.009\,8$$

$$|\delta D_1| \leqslant |\delta A| + |\delta B| \leqslant 0.065$$

$$|\mathrm{d}D_1| \leqslant 0.176\,47 \times 0.065 \approx 0.01 < 0.05 = 0.5 \times 10^{-1}$$

所以取 $D_1 = 0.2$,只准确到小数后一位。

算法 2 分成三组因子分别计算后再相乘(每组只取 6 位小数)。

$$a = \frac{0.000\,5}{0.000\,3} = 1.666\,667 (\text{有舍入})$$

$$b = \frac{0.014\,3}{0.012\,5} = 1.144\,000$$

$$c = \frac{0.001\,2}{0.013\,5} = 0.088\,889 (\text{有舍入})$$

$$\begin{aligned} D_2 &= a \times b \times c \\ &= 1.666\,667 \times 1.144\,000 \times 0.088\,889 \\ &= 1.906\,667 \times 0.088\,889 \\ &= 0.169\,482 \end{aligned}$$

$$|\delta a| \leqslant \frac{0.000\,000\,5}{1.666\,667} = 0.000\,000\,3$$

$$|\delta b| \leqslant \frac{0}{1.144\,000} = 0$$

$$|\delta c| \leqslant \frac{0.000\,000\,5}{0.088\,889} = 0.000\,005\,6$$

$$|\delta D_2| \leqslant |\delta a| + |\delta b| + |\delta c| \leqslant 0.000\,005\,9$$

$$|dD_2| \leqslant 0.169\,482 \times 0.000\,005\,9 \approx 10 \times 10^{-7} < 0.5 \times 10^{-5}$$

所以取 $D_2 = 0.169\,48$，$D_2$ 准确到小数后 5 位。

由本例中看出，采用不同的运算次序，得到的结果却大不相同。

**例 1.9**　试用 5 位有效数字计算 $e^{-5.5}$ 的值。

**解**　按以下台劳级数

$$\begin{cases} e^x = 1 + x + \dfrac{x^2}{2!} + \cdots + \dfrac{x^n}{n!} + \cdots \\ x = -5.5 \end{cases} \tag{1.42}$$

进行计算得表 1.1。

**表 1.1**

| $n$ | $\dfrac{x^n}{n!}$ | $\sum\limits_{k=1}^{n}\dfrac{x^k}{k!}$ | $n$ | $\dfrac{x^n}{n!}$ | $\sum\limits_{k=0}^{n}\dfrac{x^k}{k!}$ |
|---|---|---|---|---|---|
| 0 | 1 | 1 | 12 | 1.599 7 | 0.485 8 |
| 1 | −5.500 0 | −4.500 | 13 | −0.676 76 | −0.190 96 |
| 2 | 15.125 | 10.625 | 14 | 0.265 87 | 0.074 91 |
| 3 | −27.730 | −17.105 | 15 | −0.097 486 | −0.022 58 |
| 4 | 38.129 | 21.024 | 16 | 0.033 510 | 0.010 93 |
| 5 | −41.942 | −20.918 | 17 | −0.010 842 | 0.000 088 |
| 6 | 38.446 | 17.528 | 18 | 0.003 312 7 | 0.003 400 7 |
| 7 | −30.208 | −12.680 | 19 | −0.000 958 95 | 0.002 441 8 |
| 8 | 20.768 | 8.088 | 20 | 0.000 263 71 | 0.002 705 5 |
| 9 | −12.692 | −4.604 | 21 | −0.000 069 067 | 0.002 636 4 |
| 10 | 6.980 3 | 2.376 3 | 22 | 0.000 017 268 | 0.002 653 7 |
| 11 | −3.490 2 | −1.113 9 | 23 | −0.000 004 129 3 | 0.002 649 6 |

在本例中，假定 $x=-5.5$ 是精确数，则在计算 $e^x$ 展式中的各项数值后，由于截取为 5 位有效数字而引入舍入误差，它们累加起来会影响最终结果的准确度。由表 1.1 知，前 22 项之和为 0.002 636 4，而实值应为 0.004 086 77，可见结果没有一位是有效的。原因何在呢？事实上，由 $n=2$ 至 $n=9$ 共八项数值的舍入误差和已达到 $8\times(0.5\times10^{-3})<0.5\times10^{-2}$。即 $e^x$ 级数和的第三位小数及其以后的全部数字已不再是有效数位了。提高精度的办法之一是增加各项数值的有效数位，但它同时增加了计算量；另一种办法就是改变算法。例如可先计算 $x=5.5$ 时的 $e^x$ 部分级数和，然后求其倒数。计算中，各项的值仍取 5 位有效数字得表 1.2，由表值可见，$e^{5.5}$ 的前 18 项之和为

$$e^{5.5} \approx 1 + 5.5 + 15.125 + \cdots + 0.010\,842 = 244.70$$

其倒数值为

$$e^{-5.5} \approx \frac{1}{e^{5.5}} = \frac{1}{244.70} = 0.004\,086\,63 \tag{1.43}$$

数值 244.70 的舍入误差限为

$$\begin{aligned} &8\times(0.5\times10^{-3}) + 3\times(0.5\times10^{-4}) + 4\times(0.5\times10^{-5}) + 1\times(0.5\times10^{-6}) \\ &= 0.004\,2 < 0.5\times10^{-2} \end{aligned}$$

表 1.2

| $n$ | $\frac{x^n}{n!}$ | $\sum_{k=1}^{n}\frac{x^k}{k!}$ | $n$ | $\frac{x^n}{n!}$ | $\sum_{k=1}^{n}\frac{x^k}{k!}$ |
|---|---|---|---|---|---|
| 0 | 1 | 1 | 9 | 12.692 | 231.54 |
| 1 | 5.5 | 6.500 0 | 10 | 6.980 3 | 238.52 |
| 2 | 15.125 | 21.625 | 11 | 3.490 2 | 242.01 |
| 3 | 27.730 | 49.355 | 12 | 1.599 7 | 243.61 |
| 4 | 38.129 | 87.484 | 13 | 0.676 76 | 244.29 |
| 5 | 41.942 | 129.426 | 14 | 0.265 87 | 244.55 |
| 6 | 38.446 | 167.87 | 15 | 0.097 49 | 244.65 |
| 7 | 30.208 | 198.08 | 16 | 0.033 51 | 244.69 |
| 8 | 20.768 | 218.85 | 17 | 0.010 842 | 244.70 |

数值 0.004 086 63 的相对误差限为

$$|\delta| \leqslant \frac{0}{1} + \frac{0.5\times10^{-2}}{244.70} = 0.000\ 020\ 433$$

其绝对误差限为

$$|\Delta| \leqslant 0.004\ 086\ 63 \times 0.000\ 020\ 433 = 0.84\times10^{-7} < 0.5\times10^{-6}$$

所以 $e^{-5.5}$ 可取为 0.004 087。

**例 1.10** 求下述二次方程的根：

$$x^2 - 10^5 x + 1 = 0 \tag{1.44}$$

**解** 按二次方程求根公式

$$x_1 = \frac{10^5 + \sqrt{10^{10} - 4}}{2} \tag{1.45}$$

$$x_2 = \frac{10^5 - \sqrt{10^{10} - 4}}{2} \tag{1.46}$$

及 8 位浮点数计算得

$$x_1 = \frac{10^5 + 10^5}{2} = 10^5\text{(好)}$$

$$x_2 = \frac{10^5 - 10^5}{2} = 0\ \text{(错)}$$

产生错误的原因一是出现大数 $10^{10}$ 吃掉小数 4 的情况；二是在式(1.46)的分子部分出现两个相近数相减而导致有效数位的丧失，常称之为灾难性的抵消。为避免上述情况发生，可采取以下方法。

① 增加数值的有效数位至 11 位进行计算，这时大数已不再能吃掉小数，且两个相近数相减后仍能保留有一定的有效数位，所得结果为

$$x_1 = 0.999\ 999\ 999\ 90 \times 10^5\text{(正确)}$$

$$x_2 = 0.100\ 000\ 000\ 10 \times 10^{-4}\text{(正确)}$$

② 根据 $ax^2+bx+c=0$ 中 $b$ 的符号来选择求根公式，排除具有可能产生灾难性抵消的求根公式，保留另一求根公式。据此就可设计出以下计算公式

$$x_1 = -\frac{b+\mathrm{sgn}(b)\sqrt{b^2-4ac}}{2a} \tag{1.47}$$

式中
$$\mathrm{sgn}(b) = \begin{cases} 1, & b>0 \\ -1, & b<0 \end{cases}$$

另一根可由下式计算

$$x_2 = \frac{c}{ax_1} \tag{1.48}$$

在本例中，$a=c=1, b=-10^5$，按式(1.47)、式(1.48)计算得

$$x_1 = 1\times 10^5,\quad x_2 = 1\times 10^{-5}$$

**例 1.11**　试计算

$$I_n = \int_0^1 x^n e^{x-1}dx \qquad (n=0,1,2,\cdots,7) \tag{1.49}$$

的值。

**解**　*算法* 1　采用以下递推公式(用分部积分法求得)

$$I_n = 1-nI_{n-1} \tag{1.50}$$

进行计算。起始值 $I_0$ 取为

$$I_0 = \int_0^1 e^{x-1}dx = e^{x-1}\Big|_0^1 = 1-e^{-1} = 0.6321$$

它具有舍入误差限 $\Delta=|\Delta_0|=0.5\times10^{-4}$。在计算 $I_1, I_2, \cdots, I_7$ 的递推过程中，上述误差被逐次地放大为

$$1!\Delta, 2!\Delta, \cdots, 7!\Delta$$

由表 1.3 可见，后面的几个计算结果根本不可靠。

*算法* 2　采用以下公式

$$I_{n-1} = \frac{1-I_n}{n} \tag{1.51}$$

进行计算。用起始值 $I_7=0.1124$(具有舍入误差限 $0.5\times10^{-4}$)反向计算出 $I_6, I_5, \cdots, I_1, I_0$ 的数值，这时因 $I_7$ 的舍入误差在逐次计算中被削弱而获得较好的结果(见表 1.3)。

**表 1.3**

| $I_n$ | 算法 1 | 算法 2 | 真　值 |
|---|---|---|---|
| $I_0$ | 0.632 1 | 0.632 0 | 0.632 1 |
| $I_1$ | 0.367 9 | 0.368 0 | 0.367 9 |
| $I_2$ | 0.264 2 | 0.264 3 | 0.264 2 |
| $I_3$ | 0.207 4 | 0.207 3 | 0.207 3 |
| $I_4$ | 0.170 4 | 0.170 8 | 0.170 9 |
| $I_5$ | 0.148 | 0.145 5 | 0.145 5 |
| $I_6$ | 0.112 | 0.126 9 | 0.126 8 |
| $I_7$ | 0.216 | 0.112 4 | 0.112 4 |

由本例可见，不同的算法对舍入误差的作用是不同的，我们把舍入误差对计算结果影响小

的算法称为稳定的算法；否则称为不稳定的算法。显然，只有选用稳定的算法，才能获得较准确的结果。

综上可知，伴随计算过程而出现的舍入误差是不可避免的，它的大小与字长、计算公式的特性、计算顺序、舍入方法以及计算量的大小等因素有关。由于它直接影响到算法的数值稳定性，所以在数值方法的选择和设计时必须慎重地考虑这些因素的影响。

# §5 数值计算中的误差

## 5.1 数值计算中的误差种类

用数值方法解题一般要经过以下几个过程。

① 对于要解决的实际问题建立数学模型。

② 研制用于求解数学模型近似解的算法和过程，即将数学模型转化为数值计算公式和对数值计算的顺序。

③ 根据②确定的计算模型进行数值计算(手工计算或编制程序上机计算)，得到计算结果。

在上述过程中，可能产生的误差有以下几类。

### 5.1.1 模型误差

数学模型是指那些利用数学语言模拟现实而建立起来的有关量的描述。为了简化数学模型，以便于分析或计算，往往要忽略一些次要的因素；另外，用以描述事物的定律或定理是在特定条件下建立的，它与实际条件有所差别。这样就导致数学模型的准确解与实际问题的真解有所不同，它们之差称为模型误差或描述误差。这类误差难于做定量分析，在数值计算中，总是假定所研制的数学模型是合理的，对它的误差只作粗略的了解而不作深入的探究，以便为选择合宜的数值方法建立必要的依据。

### 5.1.2 观测误差

在数学模型中含有的参数及所需的原始数据，如温度、时间、速度、距离、电流、电压等，它们都是从观测或实验得到的数据。由于观测仪器、观测方法或某些偶然的客观因素均会引入相应的误差，这类误差叫做观测误差。通常根据测量工具或测试仪器的精度、测试水平，可以知道这类误差的上、下界。在大多数工程测量中，能得到 2～3 位有效数字。在比较重要的测量中，能得到 4～5 位有效数字。在最精确的物理测量中，能得到 6～8 位有效数字是很罕见的了。对于观测误差，在数值计算中，需要了解这类误差的范围，以便选择合理的数值方法与之适应。

### 5.1.3 截断误差

一个精确公式用近似公式代替时所产生的误差称为截断误差或公式误差或方法误差。例如一个台劳级数截取其前 $n$ 项近似代替时，其产生的截断误差为 $R_n(x)$。截断误差的大小直接影响计算结果的精度和计算量，是数值计算中必须考虑的一类误差。

### 5.1.4 舍入误差

由于数值计算只能对有限位字长的数值进行运算，这样必须对参与运算的数据作有限位字长的舍入处理，由此引入的误差称为舍入误差。舍入误差也是数值计算中必须考虑的一类误差。

## 5.2　模型与解

### 5.2.1　数学模型的解

数学模型的解按其输入数据的不同可以分为如图 1.2 所示的两类。

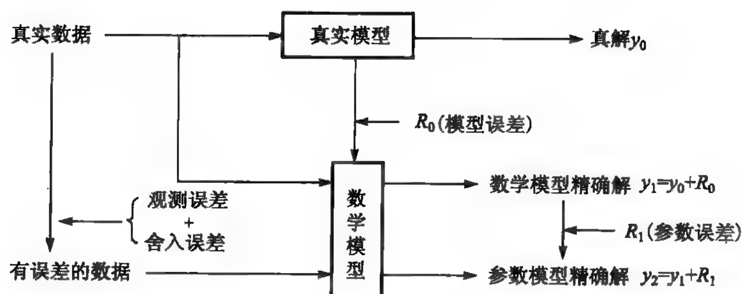



图 1.2

(1) *数学模型的精确解*

它是以真实数据输入数学模型所获得的精确解，它与真实模型的真解之差即为模型误差或描述误差。

(2) *参数模型的精确解*

它是以有误差的数据输入数学模型所获得的精确解，它与数学模型的精确解之差称为参数误差，这种误差是因数据误差而导致数学模型的精确解发生畸变所造成的。这里所述的有误差的数据指的是观测数据，其相对于真实数据的误差为观测误差；另一类有误差的数据指的是对观测数据再作舍入处理后的数据，它相对于真实数据的误差由观测误差与舍入误差之和构成。观测误差和舍入误差各自使数学模型的精确解产生相应的畸变，其畸变量之和就是参数误差，它的大小可按式(1.35)进行估计。

### 5.2.2　计算模型的解

当数学模型很复杂或不能求得其精确解时，就采用数值方法求解数学模型的数值解，这种求解模型称为计算模型，它是由一系列的计算公式(只含有加、减、乘、除运算)和计算步骤组成的。数值方法可分为精确法与近似法两种。所谓精确的数值方法指的是通过有限步精确运算就能求得数学模型精确的方法，因此其方法误差为零。相反，在近似的数值方法中，其方法误差则不为零。当使用计算模型进行求解时，对数值的运算方式可以分为精确的运算方式和近似的运算方式两种。精确的运算方式使用无限位字长的精确数进行运算，因此所有参与运算的数值，其舍入误差均为零。近似的运算方式使用有限位字长的近似数进行运算，这就形成了计算结果的舍入误差。根据输入数据、数值方法、运算方式类型的不同组合，可形成具有不同误差的数值解 $y_3 \sim y_{10}$，如图 1.3 所示(其中$\in$为舍入误差)。通常把计算模型精确运算结果和近似运算结果分别称为计算模型的精确解和计算模型的近似解。

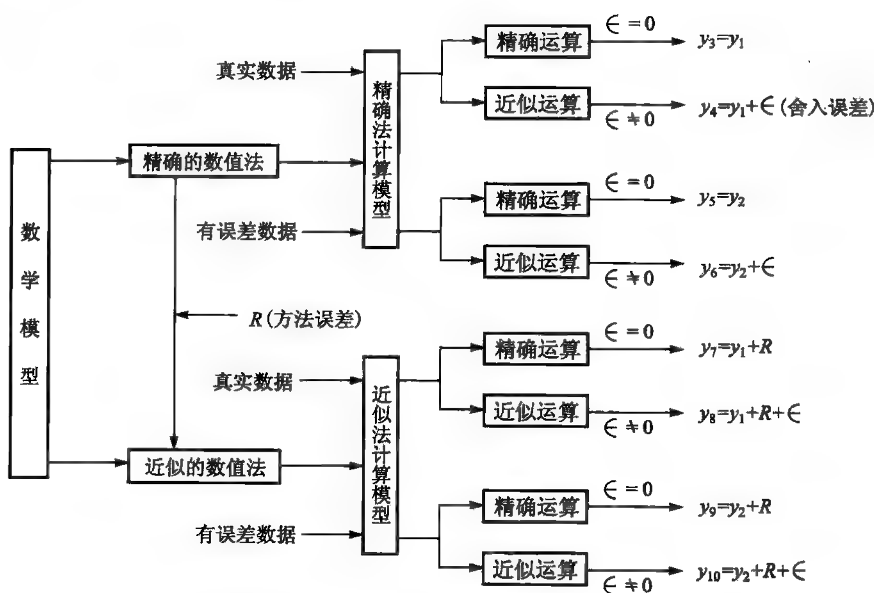



图 1.3

## 5.3 数学问题的适定性与性态

一个数学模型中的数学问题可以形式地抽象为映射 $Y=f(X)$，这里 $X$ 和 $Y$ 可以是数、多维数组(向量)、矩阵或函数。笼统地称 $X$ 为原始数据，它的映像为 $Y$。设 $D$ 为原始数据 $X$ 的定义域，根据原始数据对数学模型解的影响程度，可以将数学问题区分为适定和不适定的数学问题。在适定的情况下，其数学模型满足以下条件。

① 对 $X\in D$，数学模型的解存在且唯一。

② 满足连续性条件，即 $\|\Delta X\|\to 0$ 时，对应有 $\|\Delta Y\|\to 0$ 成立[1]。

反之，若数学模型的解多于一个；或者解不连续依赖于原始数据，则称该数学问题是不适定的。对于不适定的数学问题，如果它的解多于一个，应用数值方法求解时，就会产生求出的解是哪一个解或是否是所需解的问题。通常可用补充条件的办法使解唯一。例如，求方程根时，可先将其多个根逐一隔离开来，则在含根的每个子区域内其根唯一。如果条件②不满足，则会出现不同的 $\|\Delta X\|$ (不管它如何小)对应完全不同的数学模型解 $f(X)$的情况，这时采用数值方法求得的解可能毫无意义。因此，在不适定的情况下，必须重新研制适定性的数学模型或改善模型的适定程度，再进行数值求解。

适定的数学问题，其数学模型一般满足广义的李普希兹条件，即存在常数 $L>0,\varepsilon>0$，使对所有满足 $\|\Delta X\|<\varepsilon$ 时，有

$$\|f(X+\Delta X)-f(X)\|\leqslant L\|\Delta X\|$$

成立。若 $L$ 非常小，则对于原始数据很小的变化，数学模型解的变化也很小，就称该数学问题是“良态”问题；反之，若数学模型解的变化很大，就称为“病态”问题。

[1] 记号 $\|\cdot\|$ 表示范数，其定义参看第四章。

在病态问题中，由于观测误差或舍入误差使原始数据有微小变化时，就会导致很大的参数误差，因而原始数据的误差对数学模型的精确解有严重的影响。例如微分方程初值问题

$$y''-y=0,\quad y(0)=1,\quad y'(0)=-1 \tag{1.52}$$

该问题的解为

$$y=e^{-x} \tag{1.53}$$

如果初始条件略有误差，比如 $y(0)=1+\varepsilon, y'(0)=-1$，则相应的解为

$$y=\frac{\varepsilon}{2}e^{x}+(1+\frac{\varepsilon}{2})e^{-x} \tag{1.54}$$

它与式(1.53)相去甚远，解的性态根本不同。

这里“很大”与“很小”均系相对而言，并无数量上的严格界限。另外需要指出，数学问题的性态是针对数学问题而言的；而数值稳定性是针对数值方法而言的。数学问题的性态完全取决于该数学问题的属性，与数值方法无关。在使用数值方法求解适定的数学问题时，由性态与数值稳定性导致的误差对数值解的影响如下所示：

$$\text{数值方法}\begin{cases}\text{稳定的数值方法}\in\downarrow\begin{cases}\text{良态问题 }R_1\ \downarrow\\ \text{病态问题 }R_1\ \uparrow\uparrow\end{cases}\\ \text{不稳定的数值}\in\uparrow\uparrow\begin{cases}\text{良态问题 }R_1\ \downarrow\\ \text{病态问题 }R_1\ \uparrow\uparrow\end{cases}\end{cases}$$

综上，很容易看出以下几点。

① 用一个稳定的数值方法解一个良态问题时，在数值解中含有的 $R_1$、$\in$ 必定很小，如果数值方法的方法误差 $R$ 亦很小，则可获得接近于参数模型精确解的最佳结果。

② 当用一个稳定的数值方法求解病态问题或用一个不稳定的数值方法求解良态问题时，前者因 $R_1$ 甚大，后者因 $\in$ 甚大，致使数值解甚大地偏离数学模型的精确解。

③ 当用一个不稳定的数值方法求解病态问题时，因 $R_1$ 与 $\in$ 均甚大，视 $R_1+\in$ 的大小，数值解可能更加偏离或接近于数学模型的精确解。

为了衡量计算模型数值解的精度，由上可见，可采用参数模型的精确解作为测试的标准值，一般可要求计算模型数值解的精度小于等于标准值的精度。在参数模型的精确解不能获得的情况下，实用上可采用高精度计算模型的高精度计算结果作为该类计算模型的标准值。如果采用真解为标准值，由上面的分析可知，计算模型的数值解相对于真解的误差由 $R_0+R_1+R+\in$ 组成，当 $R_0$、$R_1$、$R$、$\in$ 的精度配置不合理时，可能导致高计算量、低精度数值解的不良结果。因此在一个计算模型的设计中，要全面地综合考虑各种因素对误差的影响并合理地配置误差的大小，才能达到以最小的计算量获取高精度数值解的理想目标。

今后我们所讨论的数学问题只限于适定的数学问题，其中的良态问题比较适宜数值求解。对于病态问题的求解，不是一般稳定算法所能解决的，必须使用特殊的算法来处理。

# §6　误差分配原则与处理方法

## 6.1　误差配置原理

如前所述，计算模型总是以参数模型为依据，以数值方法为手段而研制的求解模型，计算

模型的近似解相对于参数模型精确解的总误差(ε)由截断误差(R)和舍入误差(∈)两部分构成,可表为

$$\varepsilon = R + \in \tag{1.55}$$

现分析按数值公式计算时,ε、R、∈之间存在的几种可能的情况。

① 当∈>R时,即舍入误差较大而截断误差较小的情况。这时总误差的主部取决于舍入误差的主部。例如在例1.9中,取$e^{-5.5}$展式中前22项和为数值公式时的截断误差为

$$|R_{21}| = \left|\frac{(-5.5)^{22}}{22!}e^{-5.5\theta}\right| \leqslant \frac{(5.5)^{22}}{22!} = 0.000\,017 \qquad (0<\theta<1)$$

而$\in \approx 0.5\times10^{-2}$,属∈>R的情况。显见,在这种情况下,数值公式中数值小于∈的若干项的计算工作量是无意义的。

② 当∈<R时,这时总误差的主部取决于截断误差的主部,在这种情况下,过多位字长部分的计算工作量是无意义的。

③ 当∈≈R时,在这种情况下,不会出现过多位字长和过多项部分计算量上的浪费现象。因此,对于R、∈两类误差最为合理的配置原则是

$$R = \in \tag{1.56}$$

## 6.2 按R=∈配置的有关处理方法

针对配置原则式(1.56),对下列四种情况论述其处理方法。

(1) 给定运算误差∈,确定参与运算的数值的字长

若数值公式为

$$u = f(x_1, x_2, \cdots, x_n) \tag{1.57}$$

式中,$x_1, x_2, \cdots, x_n$为参与运算的数值,设其允许的舍入误差为$|\Delta_i| = \Delta (i=1,2,\cdots,n)$,则计算结果的舍入误差限可近似估计为

$$|du| \leqslant \sum_{i=1}^{n}\left|\frac{\partial f}{\partial x_i}\right| \cdot |\Delta_i| = \left(\sum_{i=1}^{n}\left|\frac{\partial f}{\partial x_i}\right|\right)\cdot \Delta \tag{1.58}$$

令$\left(\sum_{i=1}^{n}\left|\frac{\partial f}{\partial x_i}\right|\right)\cdot \Delta \leqslant \in$,可解得

$$\Delta \leqslant \frac{\in}{\sum_{i=1}^{n}\left|\frac{\partial f}{\partial x_i}\right|} \tag{1.59}$$

按式(1.59)的要求,就可确定出数值的字长。

**例1.12** 长方形面积$S=ab$,其中$a\approx5$ m,$b\approx200$ m,要求计算S的运算误差$\Delta S\leqslant 1\ \mathrm{m}^2$,试确定两直角边的允许误差。

**解** 令$|da| = |db| = \Delta$,因$|ds| \leqslant (|a|+|b|)\cdot\Delta$,令$(|a|+|b|)\cdot\Delta \leqslant 1\ \mathrm{m}^2$,解得

$$\Delta \leqslant \frac{1\ \mathrm{m}^2}{(5+200)\mathrm{m}^2} = \frac{1}{205} \approx 0.004\,8$$

因为0.000 5<0.004 8,所以可取$\Delta = 0.000\,5 = 0.5\times10^{-3}$,即a,b的数值至少应取至小数后3位。

(2) 数值公式已定、数值字长待定的情况

在这种情况下,能估计出该数值公式截断误差R的大小,然后令∈=R,按式(1.59)就可

确定出参与运算的数值的字长。

(3) *在总误差 $\varepsilon$ 给定的情况下，要求确定数值公式的项数和数值的字长*

在这种情况下，按误差配置的原则，我们应取

$$R=\in=\frac{\varepsilon}{2} \tag{1.60}$$

以下可按不等式

$$|R|\leqslant\varepsilon/2 \tag{1.61}$$

确定出数值公式的项数。在数值公式已定和$\in$已定的情况下，便可确定出数值的字长。

**例 1.13**　求 $y=\frac{1}{1^5}+\frac{1}{2^5}+\cdots+\frac{1}{n^5}+\cdots$之值，总误差要求为 $\varepsilon=0.001$。

**解**　在本例中，取级数的部分和

$$S_n=\sum_{k=1}^{n}\frac{1}{k^5} \tag{1.62}$$

作为数值公式，其截断误差为

$$R_n=\sum_{k=n+1}^{\infty}\frac{1}{k^5} \tag{1.63}$$

按式(1.60)，取

$$R_n=\frac{\varepsilon}{2}=0.000\,5,\quad \in=\frac{\varepsilon}{2}=0.000\,5 \tag{1.64}$$

对于式(1.63)有以下估计公式(如图 1.4 所示)

$$R_n=\sum_{k=n+1}^{\infty}\frac{1}{k^5}\leqslant\int_n^{\infty}\frac{\mathrm{d}x}{x^5}=\frac{1}{4n^4} \tag{1.65}$$

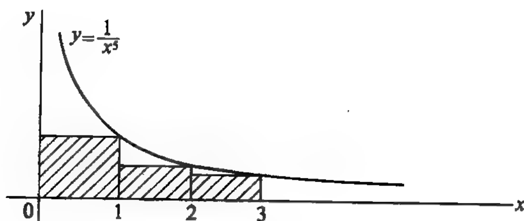



图 1.4

令

$$R_n\leqslant\frac{1}{4n^4}\leqslant 0.000\,5$$

则 $n$ 应满足

$$n\geqslant\sqrt[4]{500}\approx 4.7$$

取 $n=5$ 得计算公式

$$S_5=\sum_{k=1}^{5}\frac{1}{k^5} \tag{1.66}$$

设计算式(1.66)中的每项数值的舍入误差限为$|\Delta|$，则可按下式

$$\in=4|\Delta|\leqslant 0.000\,5$$

确定$|\Delta|$

$$|\Delta| \leqslant \frac{0.000\,5}{4} = 0.000\,125$$

可取$|\Delta| = 0.000\,05 = 0.5 \times 10^{-4} < 0.000\,125$，采用四位小数计算式(1.66)得

$$S_5 = 1 + 0.031\,3 + 0.004\,1 + 0.001\,0 + 0.000\,3 = 1.036\,7$$

再按式(1.64)的$\in = 0.5 \times 10^{-3}$ 将$S_5$舍成三位小数得$S_5 = 1.037$。

(4) 数值的字长已定、待定数值公式的项数情况

在这种情况下，可采用尝试法确定之。这时可初步选定该数值公式的项数，估计出相应的$R$、$\in$值，如果$R$与$\in$比较相近，则上述所定的项数即为所要求的项数；否则，应另选项数，重复上面的过程，直至选出符合要求的项数为止。

**例 1.14** 对于$0 \leqslant x < 1$，计算$e^x$时，利用$e^x$的台劳展开式的部分和

$$U = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} \tag{1.67}$$

近似$e^x$，若在计算中，$x$的数值及式(1.67)中各项结果均截取至小数后5位，试确定式(1.67)中应取几项为好？

**解** 因公式(1.67)的截断误差有以下估计

$$R_n(x) = \frac{e^{\theta x}}{(n+1)!}x^{n+1} < \frac{3}{(n+1)!} \quad (0 < \theta < 1) \tag{1.68}$$

初取$n=3$，按式(1.68)有

$$R_3(x) < \frac{3}{4!} = 0.125$$

$$\in_3 = 3 \times (0.5 \times 10^{-5}) = 0.000\,015$$

这时$R_3(x) > \in_3$，再取$n=6$，则有

$$R_6(x) < \frac{3}{7!} = 0.000\,59$$

$$\in_6 = 6 \times (0.5 \times 10^{-5}) = 0.000\,03$$

这时$R_6(x) > \in_6$，增大$n$至$n=7$，则有

$$R_7(x) = \frac{3}{8!} = 0.000\,074$$

$$\in_7 = 7 \times (0.5 \times 10^{-5}) = 0.000\,035$$

这时，$R_7(x) \approx \in_7$，所以在式(1.67)中应取8项进行计算为宜。

在许多使用截断误差限进行精度控制的计算场合，当数值精度低于上述预定精度或舍入误差超过上述预定精度时，则这种控制就失效了。目前对舍入误差的估计都是按最坏的情况来估算的，实际上未必是这样。由于舍入误差估计的困难性，又为了弥补可能的精度损失，粗略的做法是，视影响舍入误差诸因素的情况，按照对截断误差的预定精度要求，用比预定精度位多若干位(1～3位)的数值进行计算，最后把结果舍入到所要求的精度位上。

# 习 题 一

1.1　下列各数都是对精确数进行四舍五入得到的近似值,试分别指出它们的绝对误差限、相对误差限以及有效数字的位数。

$$a_1=0.031\,5,\quad a_2=0.301\,5,\quad a_3=31.50,\quad a_4=5\,000,\quad a_5=5\times10^3$$

1.2　已知$\sqrt{3}=1.732\,050\,808\cdots$,试写出具有三位、四位及五位有效数字的近似值,并求出它们的绝对误差限及相对误差限。

1.3　下列各近似值的绝对误差限都是 $0.5\times10^{-3}$,$a=-1.000\,31$,$b=0.042$,$c=-0.000\,32$,试指出它们有几位有效数字。

1.4　要使$\sqrt{10}$的近似值的相对误差小于 0.1%,至少要取几位有效数字?

1.5　下面计算 $y$ 的公式哪一个算得准?为什么?

(1) 已知$|x|\ll1$

(A) $y=\dfrac{1}{1+2x}-\dfrac{1-x}{1+x}$; (B) $y=\dfrac{2x^2}{(1+2x)(1+x)}$

(2) 已知$|x|\gg1$

(A) $y=\dfrac{2}{x\left(\sqrt{x+\frac{1}{x}}+\sqrt{x-\frac{1}{x}}\right)}$; (B) $y=\sqrt{x+\dfrac{1}{x}}-\sqrt{x-\dfrac{1}{x}}$

(3) 已知$|x|\ll1$

(A) $y=2\sin^2 x/x$; (B) $y=(1-\cos 2x)/x$

1.6　计算 $f=(\sqrt{2}-1)^6$,取$\sqrt{2}\approx1.4$,利用下列公式计算,哪一个得到的结果较好?

(1) $f=\dfrac{1}{(\sqrt{2}+1)^6}$　　(2) $f=(3-2\sqrt{2})^3$

(3) $f=\dfrac{1}{(3+2\sqrt{2})^3}$　　(4) $f=99-70\sqrt{2}$

1.7　下列各题怎样计算才合理?

(1) $1-\cos1^\circ$(用四位函数表求三角函数值)

(2) $\displaystyle\int_N^{N+1}\frac{\mathrm{d}x}{1+x^2}$(其中 $N$ 充分大)

(3) $\dfrac{1-\cos x}{\sin x}$(其中$|x|$充分小)

1.8　先化简下式,再求其值。

$$S=\sum_{j=2}^{1000}\frac{1}{j^2-1}$$

1.9　求$\dfrac{1}{662}-\dfrac{1}{663}$的数值。

1.10　求方程 $x^2-56x+1=0$ 的两个根,使它至少具有四位有效数字。

1.11　给定方程组

$$\begin{cases} 3x + ay = 10 \\ 5x + by = 20 \end{cases}$$

其中 $a=2.100\pm5\times10^{-4}$，$b=3.300\pm5\times10^{-4}$，试估计计算 $x,y$ 的误差范围。

1.12 $\sin x=x-\frac{x^3}{3!}+\frac{x^5}{5!}\cdots$，试问用台劳级数计算 sin1，如果要有 8 位有效数字，需要在级数中取几项？并且运算的数要有多少位才能达到要求？

1.13 假如有一种算法求$\sqrt{a}$可得到 6 位有效数字，问为了使$\sqrt{\pi}$有 4 位有效数字，$\pi$ 应取几位有效数字？

# 第二章　方程(组)的迭代解法

## §1 引　言

在数学问题的求解过程中，经常遇到求解函数方程

$$f(x)=0 \tag{2.1}$$

的问题。这里，$f(x)$是单变量 $x$ 的函数，它可以是代数多项式

$$f(x)=a_0x^n+a_1x^{n-1}+\cdots+a_{n-1}x+a_n \qquad (a_0\neq 0)$$

也可以是超越函数，如三角函数、指数函数的复合函数等。满足方程(2.1)的 $x$ 值通常叫做方程的根或解，也叫函数 $f(x)$的零点。如果 $f(x)=(x-\alpha)^m g(x)$且 $g(\alpha)\neq 0$，则称 $\alpha$ 为 $f(x)=0$ 的 $m$ 重根。$m=1$ 为单根，$m>1$ 为重根。

由于多项式是一类非常特殊的函数，人们设计了算法来求它的零点。从挖掘出的古巴比伦人的泥版文件中发现，远在公元前 1700 年的古巴比伦人就已有关于一、二次方程的解法。以后人类又经三千多年的求索，至 1535 年意大利数学家坦特格里亚(Tor Taglia，1499—1557 年)发现了三次方程的解法，卡当(H. Cardano，1501—1570 年)从他那里得到了这种解法，于 1545 年在名著《大法》中公布了三次方程的公式解，名为卡当算法。后来卡当的学生弗瑞里(Ferrari，1522—1565 年)又提出了四次方程的解法。此成果更激发了数学家们的兴奋情绪，但在以后的两个世纪中，求索工作始终没有成效，导致人们对高次代数方程解的存在性产生怀疑。至 1799 年，高斯证明了代数方程必有一个实根或复根的定理，称此为代数基本定理，并由之可以立刻推知 $n$ 次代数方程必有 $n$ 个实根或复根。但在以后的几十年中，仍然没有找出高次代数方程的公式解。一直到 18 世纪，法国数学家拉格朗日用根置换方法统一了二、三、四次代数方程的解法，然而当他用这个理顺了的解法和理论去求解五次代数方程时却碰了钉子而未能如愿，才冷静地思考：想找而又找不到的东西是否根本就不存在？开始意识到可能有潜藏于其中的奥秘，用现代术语表示，那就是置换群理论问题。

在探索五次以上代数方程解的艰难历程中，第一个有重大突破的是挪威数学家阿贝尔(N. Abel，1802—1829 年)，1824 年年轻的阿贝尔发表了“五次方程代数解法不可能存在”的论文，但并未受到重视，连数学大师高斯也未理解这项成就的重要意义。在贫困中度过 27 个春秋的阿贝尔因结核病死于 1829 年。1828 年名不见经传的 17 岁法国数学家伽罗华(E. Galois，1811—1832 年)写出了划时代的论文“关于五次方程的代数解法问题”，指出即使在公式中容许用 $n$ 次方根，并用类似的算法求五次或更高次代数方程的根是不可能的。文章虽经柯西呈交至法兰西科学院，因其辈分太低遭到冷遇且文稿丢失。1830 年伽罗华再进科学院递稿，经泊松院士判审为“完全不能理解”而终结。伽罗华死于 1832 年，死前，他把关于五次代数方程求解的研究成果写成长信，留给了人类。14 年后，法国数学家刘维尔(J. Liouville，1809—1882 年)整理并发表了伽罗华的遗作，人们才意识到这项近代数学发展史上的重要成果是何等的宝贵，深为伽罗华之死而惋惜。38 年后，即 1870 年，法国数学家若当(C. Jordan，1838—1922 年)在专著《论置换与代数方程》中阐述了伽罗华的思想，一门现代代数的分支——群论诞生了。

此后，人们并未终止对代数方程求解算法的研究工作，在前几个世纪中，曾开发出一些求解代数方程的有效算法，它们构成了数值分析中的古典算法，如 Honer，Graeffe，Bernoulli 等算法。而在计算机时代，人们熟知的有 Rutishauser，Lehmer，Lin，Bairstow，Bareiss 算法及其他许多算法。至于超越方程则不存在一般的求根公式。

本章介绍求方程实根的迭代解法，它既可以用来求解代数方程，也可以用来求解超越方程。运用迭代法求解方程的根应解决以下两个问题：确定根的初值 $x_0$；将 $x_0$ 进一步精确化到所需要的精度。

# §2 迭代解法

设方程(2.1)的根为 $\alpha$，则 $f(\alpha)\equiv 0$。为求 $\alpha$ 的数值，可初取一个与 $\alpha$ 接近的初始近似值 $x_0$，然后通过计算出 $f(x_0)$ 的数值并测试其是否为零，若等于零，则 $x_0$ 就是所求的根 $\alpha$；在不等于零的情况下，可改变 $x_0$ 的数值(其改变量设为 $\Delta x_0$)为新的一次近似值 $x_1=x_0+\Delta x_0$，再重复以上过程，直至新的近似值满足方程为止。以上不断尝试的计算过程称为迭代过程，它本质上是一种逐次逼近根 $\alpha$ 的方法，称为迭代解法，可用图 2.1 表示。

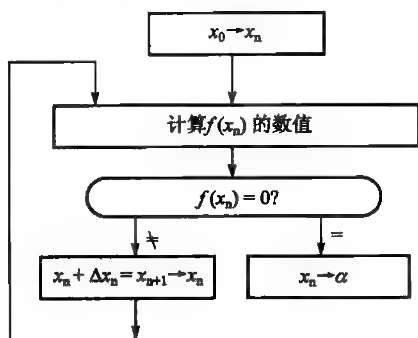



图 2.1 迭代解法

从上可见，迭代解法中的一个关键问题就是如何确定 $n$ 次近似值 $x_n$ 的修改量 $\Delta x_n$ 的问题。目前已有许多计算 $\Delta x_n$ 的不同算法，相应地形成了不同的迭代解法。为了能够直接地检测出 $x_n$ 的修改量 $\Delta x_n$，可将 $f(x)=0$ 作 $x=\varphi(x)$ 形式的等价分解，利用 $x_n$ 计算出 $\varphi(x_n)$ 的数值，就能发现方程右端值 $\varphi(x_n)$ 与左端值 $x_n$ 间的差值，不妨取定此差值为 $\Delta x_n$

$$\Delta x_n=\varphi(x_n)-x_n \tag{2.2}$$

则新近似值 $x_{n+1}$ 的计算公式(称为迭代公式)为

$$x_{n+1}=x_n+[\varphi(x_n)-x_n]=\varphi(x_n),n=0,1,2,\cdots \tag{2.3}$$

这时图 2.1 转化为图 2.2。在这种迭代法中，它将图 2.1 中计算 $f(x_n)$ 的过程与确定 $x_{n+1}$ 的修改过程融为一体，使整个迭代过程得到了简化，一般称为简单迭代法。当 $|\Delta x_n|<\varepsilon$(精度要求)时，可终止迭代并取 $x_{n+1}=\alpha$。

不难发现，由式(2.2)取定的 $\Delta x_n$ 就是方程 $x=\varphi(x)$ 当 $x=x_n$ 时的右、左两边数值之差，数学上定义它为残差(或残余)表示为

$$R_n=\varphi(x_n)-x_n \tag{2.4}$$

更一般地，选取残差的适当量作为 $\Delta x_n$ 可能获得更好的修改效果，即可取

$$\Delta x_n=wR_n \tag{2.5}$$

其中 $w$ 为对 $R_n$ 引入的系数值，相应的迭代公式为

$$x_{n+1}=x_n+wR_n,n=0,1,2,\cdots \tag{2.6}$$

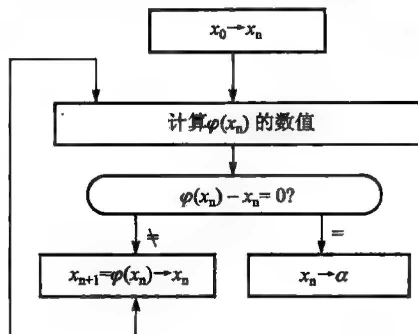



图 2.2 简单迭代法

特别当 $w=1$ 时,式(2.6)与式(2.3)一致。概括起来,迭代解法就是使用方程的残差通过迭代过程求取方程根的方法,下面具体叙述简单迭代法(简称迭代法)的求解问题。

## 2.1　根的初值确定方法

为了确定根的初值,首先必须圈定根所界的范围,称为圈定根或根的隔离。在上述基础上,采取适当的数值方法确定出具有一定精度要求的初值。对于代数方程,其根的个数(实的或复的)与其次数相同,并且已有许多关于圈定根的范围及确定所在范围内根的个数的定理和有效方法,都可取用。至于超越方程,其根可能是一个、几个或无解,且没有什么固定的圈根方法。

求方程根的问题,就几何上讲,就是求曲线 $y=f(x)$ 与 $x$ 轴交点的横坐标,作为一个根位于某个子区间内的标志常依据下面的定理。

**定理 2.1**　设 $f(x)$ 为区间 $[a,b]$ 上的连续函数,如果

$$f(a)f(b)<0$$

则 $[a,b]$ 内至少有一个实根。如果 $f(x)$ 在 $[a,b]$ 上还是单调函数,则仅有一个实根。

运用上面定理,就可确定根所在的子区间,方法有以下三种。

### 2.1.1　画图法

此法通过画出 $y=f(x)$ 的略图,从而确定出曲线与 $x$ 轴交点的大致位置。为使画图简便,可将 $f(x)=0$ 分解成 $\varphi_1(x)=\varphi_2(x)$ 的形式,其中 $\varphi_1(x)$ 与 $\varphi_2(x)$ 是较易画出其图形的函数。这时 $\varphi_1(x)$ 与 $\varphi_2(x)$ 两曲线交点的横坐标所在的子区间即为含根区间。例如

$$x\lg x-1=0$$

可以改写为

$$\lg x=\frac{1}{x}$$

再分别画出对数曲线 $y=\lg x$ 与双曲线 $y=\frac{1}{x}$,它们交点的横坐标位于区间[2,3]内,如图 2.3 所示。

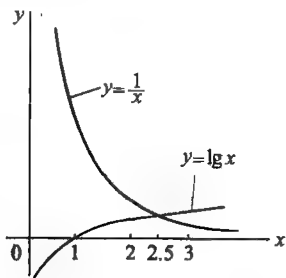



图 2.3

### 2.1.2　扫描法

对于给定的 $f(x)$ 及含根区间 $[A,B]$,从 $x_0=A$ 出发,以步长 $h=\frac{B-A}{n}$($n$ 为正整数),在 $[A,B]$ 内取定节点:$x_i=x_0+ih$ $(i=0,1,2,\cdots,n)$。然后从左至右检查 $f(x_i)$ 的符号,如发现 $f(x_k)$ 与 $f(x_{k-1})$ 异号,则得到一个有根子区间 $[x_{k-1}, x_k]$,同法继续下去,就可在 $[A,B]$ 内找出包含根的全部子区间。

用这种逐步搜索法进行实根隔离的关键是选取合适的步长。如果 $h$ 过大,则可能遗漏根;如果 $h$ 太小,虽可得到较小的有根子区间,但工作量较大。

### 2.1.3　对分法

为了获得指定精度要求的初值,可在以上隔离根的基础上采用对分法进一步缩小含根子区间,方法如下。

设 $[x_{k-1}, x_k]$ 为含根子区间,初值相对于根的误差要求为 $\varepsilon$,令 $a=x_{k-1}$,$b=x_k$,计算 $f(a)$,$f(b)$ 后进行如下计算:

① 取$[a,b]$的中点$r=\frac{a+b}{2}$,计算$f(r)$。

② 若$f(a)\cdot f(r)>0$,则取$a=r$;否则取$b=r$。

③ 若$b-a>\varepsilon$,转向①;否则结束。

经$n$次对分后获得如下含根区间

$$[a,b],[a_1,b_1],[a_2,b_2],\cdots,[a_n,b_n]$$

其区间长度逐次减半,设经$n$次对分后,其区间长度已小于$\varepsilon$,则有以下估计$n$的公式

$$\frac{b-a}{2^n}\leqslant\varepsilon$$

$$n\geqslant\frac{\ln(b-a)-\ln\varepsilon}{\ln 2} \tag{2.7}$$

这时我们可在$[a_n,b_n]$内任意取定一点$x_0$作为初值。

**例 2.1** 用对分法求方程

$$f(x)=x^3-x-1=0$$

在区间[1,1.5]内的根的初值,要求相对于根的精度准确到小数后第二位($\varepsilon=0.5\times10^{-2}$)。

**解** 这里$a=1,b=1.5,f(1)<0,f(1.5)>0$。对分计算如下:

| $i$ | $[a_i,b_i]$ | $r_i=\frac{a_i+b_i}{2}$ | $f(r_i)$ |
|---|---|---|---|
| 0 | [1,1.5] | 1.25 | $f(1.25)<0$ |
| 1 | [1.25,1.5] | 1.375 | $f(1.375)>0$ |
| 2 | [1.25,1.375] | 1.312 5 | $f(1.3125)<0$ |
| 3 | [1.312 5,1.375] | 1.343 75 | $f(1.34375)>0$ |
| 4 | [1.312 5,1.343 75] | 1.328 1 | $f(1.3281)>0$ |
| 5 | [1.312 5,1.328 1] | 1.320 3 | $f(1.3203)<0$ |
| 6 | [1.320 3,1.328 1] | 1.324 2 | $f(1.3242)>0$ |
| 7 | [1.320 3,1.324 2] | | |

这时区间[1.320 3,1.324 2]的长度为0.003 9<0.005,已满足精度要求,可在该区间内任取一点作为初值$x_0$,比如取初值为

$$x_0=\frac{1.3203+1.3242}{2}=1.3223$$

对分法同样可用来求取所需精度的方程的根,但计算工作量较大。

## 2.2 迭代法的求解过程

迭代法的求解过程分以下两步进行。

### 2.2.1 建立迭代公式

它是由方程$f(x)=0$出发,将其分解为下列等价形式

$$x=\varphi(x)$$

式中,$\varphi(x)$叫做方程的迭代函数。对于具体的方程,可以用不同的方法进行分解,例如

$$f(x)=x^3+2x^2-4=0 \tag{2.8}$$

可以分解成

$$\begin{cases} x=x+f(x)=x^3+2x^2+x-4 \\ x=\left[2\left(\dfrac{2}{x}-x\right)\right]^{\frac{1}{2}} \\ x=\left(2-\dfrac{x^3}{2}\right)^{\frac{1}{2}} \\ x=2\left(\dfrac{1}{2+x}\right)^{\frac{1}{2}} \\ x=x-\dfrac{x^3+2x^2-4}{3x^2+4x} \end{cases} \tag{2.9}$$

等等，当 $f(x)=0$ 难于作 $x=\varphi(x)$ 分解时，可在 $f(x)=0$ 两边附加 $x$ 后得迭代函数 $x=x+f(x)=\varphi(x)$。

### 2.2.2　迭代计算

由初值 $x_0$ 出发，按式(2.3)进行计算如下

$$x_1=\varphi(x_0),\quad x_2=\varphi(x_1),\quad x_3=\varphi(x_2),\cdots \tag{2.10}$$

或简记为

$$x_{n+1}=\varphi(x_n)\qquad (n=0,1,2,\cdots) \tag{2.11}$$

称公式(2.11)中形成的 $x_0,x_1,x_2,\cdots$ 为迭代序列，相应的值称为根的零次、一次、二次……近似值，序列的计算过程式(2.10)称为迭代过程。如果序列 $x_0,x_1,x_2,\cdots$ 收敛于 $\alpha$，即：

$$\lim_{n\to\infty}x_n=\alpha \tag{2.12}$$

则 $\alpha$ 就是方程的根。事实上，对式(2.11)两边取极限得

$$\lim_{n\to\infty}x_{n+1}=\lim_{n\to\infty}\varphi(x_n)=\varphi(\lim_{n\to\infty}x_n)$$

$$\alpha=\varphi(\alpha) \tag{2.13}$$

证得 $\alpha$ 是方程 $x=\varphi(x)$ 的根，亦即方程 $f(x)=0$ 的根。

以上求解方法称为方程的迭代解法或简单迭代法。

**例 2.2**　求 $xe^x-1=0$ 的根，要求根的近似值稳定至小数后 5 位。

**解**　我们将 $xe^x-1=0$ 分解为

$$x=e^{-x} \tag{2.14}$$

图形 $y=x$ 与 $y=e^{-x}$ 交点的横坐标为 $\alpha\in[0.5,0.6]$。今取 $x_0=0.5$，按迭代公式

$$x_{n+1}=e^{-x_n} \tag{2.15}$$

计算得表 2.1。经 17 次迭代后得方程的根为 0.567 14。

**表 2.1**

| $n$ | $x_n$ | $n$ | $x_n$ | $n$ | $x_n$ |
|---|---|---|---|---|---|
| 0 | 0.5 | 6 | 0.564 86 | 12 | 0.567 07 |
| 1 | 0.606 53 | 7 | 0.568 44 | 13 | 0.567 18 |
| 2 | 0.545 24 | 8 | 0.566 41 | 14 | 0.567 12 |
| 3 | 0.579 70 | 9 | 0.567 56 | 15 | 0.567 15 |
| 4 | 0.560 06 | 10 | 0.566 91 | 16 | 0.567 14 |
| 5 | 0.571 17 | 11 | 0.567 27 | 17 | 0.567 14 |

**例 2.3** 用迭代法求方程

$$f(x)=x^3-x-1=0 \tag{2.16}$$

在 $x_0=1.5$ 附近的根。

**解**:将式(2.16)改写为

$$x=\sqrt[3]{1+x} \tag{2.17}$$

按上式建立如下迭代公式

$$x_{n+1}=\sqrt[3]{1+x_n} \tag{2.18}$$

取 $x_0=1.5$ 逐次迭代得

$$x_0=1.5, x_1=1.357\ 21, x_2=1.330\ 86, x_3=1.325\ 88, x_4=1.132\ 494, x_5=1.324\ 76,$$
$$x_6=1.324\ 73, x_7=1.324\ 72, x_8=1.324\ 72$$

可取 $\alpha=1.324\ 72$ 为方程(2.16)的根。

如果改写方程为

$$x=x^3-1 \tag{2.19}$$

并建立迭代公式

$$x_{n+1}=x_n^3-1 \tag{2.20}$$

则得 $x_0=1.5, x_1=2.375, x_2=12.39, \cdots$。各次迭代值愈来愈大,这时迭代序列的极限不存在,在这种情况下,迭代法失效。这里,我们把迭代序列有极限的迭代过程称为收敛的迭代过程;否则就是发散的迭代过程。

## 2.3 迭代解法的几何意义

设方程 $f(x)=0$ 的根为 $\alpha$,对于根 $\alpha$ 的近似值 $x_0$ 显然有 $x_0\neq\varphi(x_0)$,否则 $x_0$ 就是根 $\alpha$ 了。今从 $x_0$ 出发,通过一次迭代得到 $x_1=\varphi(x_0)$,在曲线 $y_2(x)=\varphi(x)$ 上得到 $A_0$ 点(图 2.4)。为了从图上得到横坐标 $x_1$,可过 $A_0$ 点作 $x$ 轴的平行线交 $y_1(x)=x$ 于 $A'_0$ 点,再由 $A'_0$ 点向横轴作垂线得交点,此交点即是 $x_1$。以下仿此推导下去,便可循折线 $A_0A'_0A_1A'_1\cdots$ 趋近 $A$ 点,相应的横坐标 $x_0, x_1, x_2, \cdots$ 则从根的一侧趋近根 $\alpha$。各次迭代值也可以从根的两侧往复地趋向根 $\alpha$,如图 2.5 所示。迭代过程发散的情况如图 2.6、图 2.7 所示。其中图 2.7 中的迭代方式

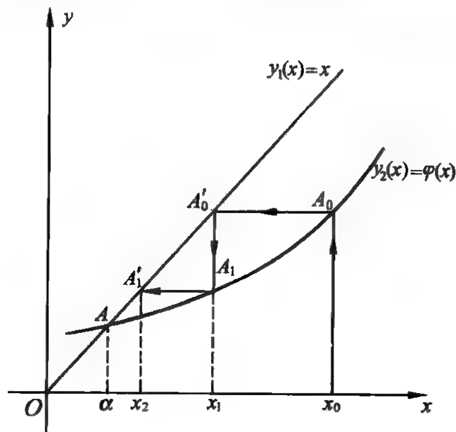



图 2.4

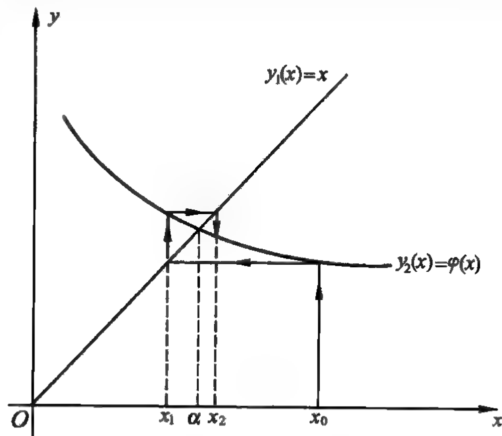



图 2.5

称为死循环,在迭代过程中反复出现两个相同的迭代值,属迭代发散过程。为防止因出现死循环而浪费计算量,可在迭代计算过程中加入最大迭代次数 $N$ 的控制,当迭代次数大于 $N$ 时,就认为无效而终止迭代计算。

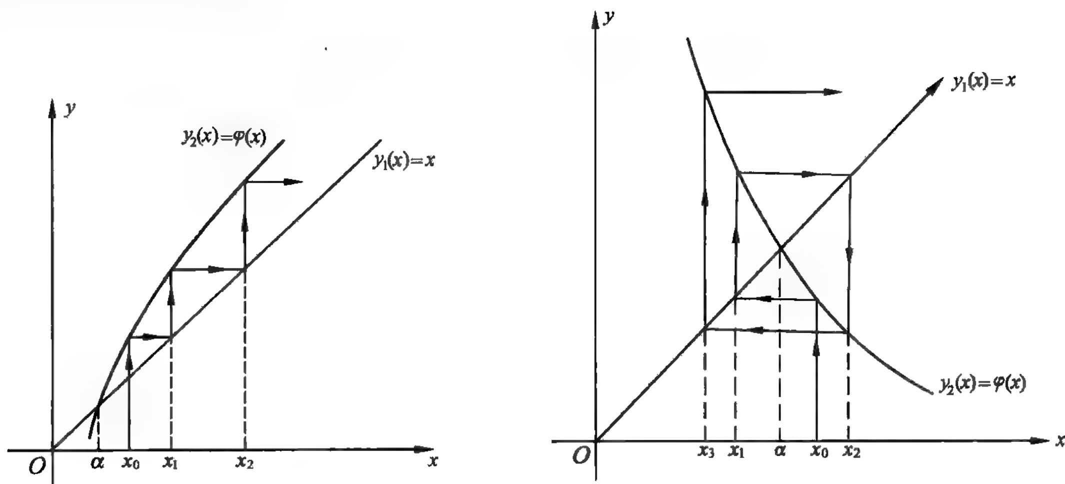



图 2.6

对于迭代法来说,一般需要讨论的基本问题是,迭代函数的构造、迭代序列的收敛性、收敛速度以及误差估计。

## 2.4　迭代法的收敛性

一个迭代法的迭代序列是否收敛不仅取决于迭代函数在根附近的性质,还与初值的选取范围有关。若从任何可取的初值出发都能保证收敛,则称之为大范围收敛。若为了保证收敛性,必须选取初值充分接近于所要求的根,则称之为局部收敛。通常,局部收敛方法比大范围收敛方法收敛得快。因此,一个合理的算法是先用一种大范围收敛方法求得接近于根的近似值(如对分法),再将它作为初值使用局部收敛方法(如迭代法)。这里讨论迭代法的收敛性时,指的均是局部收敛性。

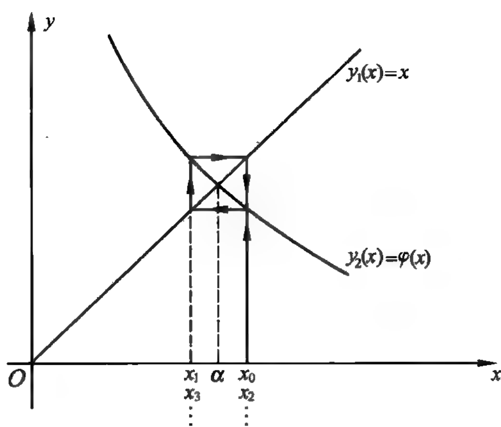



图 2.7

设 $\varphi(x)$ 在包含根的区间内具有连续的一阶导数,因 $\alpha$ 为根式(2.21)恒成立

$$\alpha=\varphi(\alpha) \tag{2.21}$$

记各次近似值的误差为 $|x_i-\alpha|$ $(i=0,1,2,\cdots)$,则有以下关系式

$$\begin{cases} |x_1-\alpha|=|\varphi(x_0)-\varphi(\alpha)|=|\varphi'(\xi_1)|\cdot|x_0-\alpha| \\ |x_2-\alpha|=|\varphi(x_1)-\varphi(\alpha)|=|\varphi'(\xi_2)|\cdot|x_1-\alpha| \\ |x_3-\alpha|=|\varphi(x_2)-\varphi(\alpha)|=|\varphi'(\xi_3)|\cdot|x_2-\alpha| \\ \cdots \\ |x_{n+1}-\alpha|=|\varphi(x_n)-\varphi(\alpha)|=|\varphi'(\xi_{n+1})|\cdot|x_n-\alpha| \end{cases} \tag{2.22}$$

将以上各式的两边分别相乘后便得到

$$|x_{n+1}-\alpha|=|\varphi'(\xi_1)|\cdot|\varphi'(\xi_2)|\cdots|\varphi'(\xi_{n+1})|\cdot|x_0-\alpha| \tag{2.23}$$

从上式可见,若有

$$|\varphi'(x)|\leqslant q<1, x\in[a,b] \tag{2.24}$$

成立,则必有下列不等式成立

$$|x_{n+1}-\alpha|\leqslant q^{n+1}|x_0-\alpha| \tag{2.25}$$

当 $n\to\infty$ 时,$q^{n+1}\to 0$。所以

$$\lim_{n\to\infty}|x_{n+1}-\alpha|=0,\text{即}\lim_{n\to\infty}x_{n+1}=\alpha$$

从而证得迭代过程收敛。由此可得到以下定理。

**定理 2.2** 设 $\varphi(x)$ 在包含根 $\alpha$ 的区间 $[a,b]$ 上可微,且满足条件

$$|\varphi'(x)|\leqslant q<1,\ x\in[a,b] \tag{2.26}$$

则对任何 $x_0\in[a,b]$,迭代过程(2.3)一定收敛。

条件(2.26)为迭代序列收敛的充分条件,$q$ 值为新、旧迭代值之误差比值中最大的绝对值。$q$ 愈小,则新近似值相对于旧近似值就更加接近于 $\alpha$,所以 $q$ 值的大小可以反映出一次迭代中新近似值接近于根 $\alpha$ 的快慢程度,一般认为 $q<1/10$ 收敛是比较快的;$q>1/2$ 则认为收敛是慢的。在实际应用时,当 $\varphi'(x)$ 连续且 $[a,b]$ 较小,则 $\varphi'(x)$ 在 $[a,b]$ 上的值变化不大,可采用下式

$$|\varphi'(x_0)|<1 \tag{2.27}$$

代替式(2.26)来判断迭代过程的收敛性。

类似地可以证明以下定理。

**定理 2.3** 若 $[a,b]$ 上的连续函数 $\varphi(x)$ 满足李普希兹条件

$$|\varphi(x_1)-\varphi(x_2)|\leqslant L|x_1-x_2| \qquad (0\leqslant L<1) \tag{2.28}$$

则迭代过程式(2.3)一定收敛。

为了比较迭代法收敛的快慢程度(通常称为收敛速度),我们引入如下衡量指标——收敛阶数。设迭代序列收敛,如果存在正数 $r$ 和 $c$,使得下式

$$\lim_{n\to\infty}\frac{|x_{n+1}-\alpha|}{|x_n-\alpha|^r}=c(\text{常数}) \tag{2.29}$$

成立,则称该迭代过程是 $r$ 阶收敛的或称收敛的阶为 $r$,常数 $c$ 称为渐近误差常数。式(2.29)表明,新近似值误差的绝对值与旧近似值误差绝对值的 $r$ 次方成正比。当 $|x_n-\alpha|<1$ 时,则 $r$ 愈大,$|x_n-\alpha|^r$ 的值就愈小,与之成比例变化的 $|x_{n+1}-\alpha|$ 也愈小,所以收敛就愈快。对于 $r=1$,其收敛性就取决于常数 $c$,当 $c<1$ 时,其误差依 $c$ 比例地下降,称为线性收敛的;$r>1$ 为超线性收敛的。其中 $r=2$ 称为平方收敛,在这种情况下,若 $|x_n-\alpha|\leqslant 0.5\times10^{-k}$,则就有 $|x_n-\alpha|^2\leqslant 0.25\times10^{-2k}<0.5\times10^{-2k}$,而 $|x_{n+1}-\alpha|$ 与 $|x_n-\alpha|^2$ 成正比,显见 $x_{n+1}$ 应有 $2k$ 位有效数字,它较 $x_n$ 的有效数位约增加了一倍。至于更高阶($r>2$)的迭代过程,新近似值有效数位的增加数就更加显著了。

常数 $c$ 与 $f(x)$ 在 $x=\alpha$ 附近的性态有关,它对收敛速度起一定的影响。将式(2.29)改写为

$$|x_{n+1}-\alpha|\approx|c^{\frac{1}{r}}(x_n-\alpha)|^r$$

后可以看出,如果 $c$ 的数值较大或阶 $r$ 较小使 $c^{\frac{1}{r}}$ 较大时,其收敛速度就要慢一些;相反,如果 $c$ 的数值较小且 $r$ 较大使 $c^{\frac{1}{r}}$ 较小时,其收敛速度就要高一些。此外,$c$ 值的大小还影响到收敛范围。事实上,反复应用式(2.29)可导得以下关系式

$$|x_1-\alpha|\approx c|x_0-\alpha|^r$$
$$|x_2-\alpha|\approx c|x_1-\alpha|^r=c^{r+1}|x_0-\alpha|^{r^2}$$
$$|x_3-\alpha|\approx c|x_2-\alpha|^r=c^{r^2+r+1}|x_0-\alpha|^{r^3}$$
$$\cdots$$
$$|x_n-\alpha|\approx c^{r^{n-1}+r^{n-2}+\cdots+r+1}|x_0-\alpha|^{r^n}$$
$$=c^{\frac{r^n-1}{r-1}}|x_0-\alpha|^{r^n}=c^{\left(\frac{r^n}{r-1}-\frac{1}{r-1}\right)}|x_0-\alpha|^{r^n}$$
$$=c^{-\frac{1}{r-1}}\left(c^{\frac{1}{r-1}}|x_{10}-\alpha|\right)^{r^n}$$

此处假定了 $x_0$ 已充分地接于 $\alpha$。如果选取初值使

$$c^{\frac{1}{r-1}}|x_0-\alpha|<1$$

即

$$|x_0-\alpha|<\frac{1}{c^{\frac{1}{r-1}}}$$

成立,则当 $n\to\infty$,就有 $x_n\to\alpha$。由此看出,在 $r\geqslant 2$ 时,虽然 $c$ 值不受小于 1 的限制,但当 $c$ 值越大时,就要求 $x_0$ 越接近于 $\alpha$。

不同的迭代法具有不同的收敛阶数,下面的定理提供了一个测定收敛阶的方法。

**定理 2.4**　设迭代函数 $\varphi(x)$在 $\alpha$ 邻近有 $r$ 阶连续导数,且满足

$$\varphi'(\alpha)=\varphi''(\alpha)=\cdots=\varphi^{(r-1)}(\alpha)=0 \tag{2.30}$$

及 $\varphi^{(r)}(\alpha)\neq 0$,则该迭代过程是 $r$ 阶收敛的。

**证**　因

$$x_{n+1}=\varphi(x_n)=\varphi[\alpha+(x_n-\alpha)]$$
$$=\varphi(\alpha)+(x_n-\alpha)\frac{\varphi'(\alpha)}{1!}+(x_n-\alpha)^2\frac{\varphi''(\alpha)}{2!}+\cdots+(x_n-\alpha)^{r-1}\frac{\varphi^{(r-1)}(\alpha)}{(r-1)!}+$$
$$(x_n-\alpha)^r\frac{\varphi^{(r)}(\xi)}{r!}$$
$$=\alpha+(x_n-\alpha)^r\frac{\varphi^{(r)}(\xi)}{r!}\qquad(\xi\in(\alpha,x_n))$$

于是

$$\frac{x_{n+1}-\alpha}{(x_n-\alpha)^r}=\frac{\varphi^{(r)}(\xi)}{r!}$$

当 $n\to\infty$时,$x_n\to\alpha$,故 $\xi\to\alpha$,从而有

$$\lim_{n\to\infty}\frac{|x_{n+1}-\alpha|}{|x_n-\alpha|^r}=\frac{|\varphi^{(r)}(\alpha)|}{r!}=\text{常数}(\neq 0)$$

上式表明,迭代过程是 $r$ 阶收敛的。

## 2.5　迭代序列的误差估计

迭代过程 $x_{n+1}=\varphi(x_n)(n=0,1,2,\cdots)$不应无休止地进行下去,当

$$|x_{n+1}-\alpha|<\varepsilon(\text{精度要求}) \tag{2.31}$$

满足时,应终止迭代计算,这时 $x_{n+1}$就是满足精度要求的根。因为 $\alpha$ 未知,$x_{n+1}-\alpha$ 亦属未知,因

此式(2.31)无法实际应用。为此,我们要寻找一些公式间接地来判定,这些公式均属$|x_{n+1}-\alpha|$的上界公式或称估计公式。如果上界公式之值小于ε,则必有式(2.31)成立,从而解决了迭代过程结束的判断问题。下面推导三个上界式,在推导公式之前,均假定$|\varphi'(x)|\leqslant q<1, x\in[a,b]$条件满足,其中第三个公式,增加条件$|f'(x)|\geqslant m>0$且都按精确计算获得迭代序列的数值。

上界公式1
$$|x_{n+1}-\alpha|\leqslant\frac{q}{1-q}|x_{n+1}-x_n| \tag{2.32}$$

**证** 因

$$\begin{aligned}x_{n+1}-\alpha&=\varphi(x_n)-\varphi(\alpha)\\&=-[\varphi(x_{n+1})-\varphi(x_n)]+[\varphi(x_{n+1})-\varphi(\alpha)]\\&=-\varphi'(\xi_1)(x_{n+1}-x_n)+\varphi'(\xi_2)(x_{n+1}-\alpha)\end{aligned}$$

所以有
$$[1-\varphi'(\xi_2)](x_{n+1}-\alpha)=-\varphi'(\xi_1)(x_{n+1}-x_n)$$

$$|x_{n+1}-\alpha|=\left|\frac{\varphi'(\xi_1)}{1-\varphi'(\xi_2)}\right|\cdot|x_{n+1}-x_n|\leqslant\frac{q}{1-q}|x_{n+1}-x_n|$$ (证毕)

为了使上界式(2.32)使用上简便,我们分以下两种情况来简化它。

(1) 当$0<q\leqslant\frac{1}{2}$时,$\frac{q}{1-q}\leqslant1$,这时式(2.32)可以进一步化为

$$|x_{n+1}-\alpha|\leqslant\frac{q}{1-q}|x_{n+1}-x_n|\leqslant|x_{n+1}-x_n|$$

当
$$|x_{n+1}-x_n|\leqslant\varepsilon \tag{2.33}$$
时,就能保证$|x_{n+1}-\alpha|\leqslant\varepsilon$成立。

(2) 当$q>\frac{1}{2}$时,$\frac{q}{1-q}>1$,这时只能要求

$$\frac{q}{1-q}|x_{n+1}-x_n|\leqslant\varepsilon \tag{2.34}$$

来保证$|x_{n+1}-\alpha|\leqslant\varepsilon$的要求。为了简化式(2.34),我们将它改写为等价的形式

$$|x_{n+1}-x_n|\leqslant\frac{1-q}{q}\varepsilon \tag{2.35}$$

令$\bar{\varepsilon}=\frac{1-q}{q}\varepsilon$,则式(2.35)可化为与式(2.33)相同的形式

$$|x_{n+1}-x_n|\leqslant\bar{\varepsilon} \tag{2.36}$$

式中$\bar{\varepsilon}<\varepsilon$。最后式(2.33)与式(2.36)可统一表示为

$$|x_{n+1}-x_n|\leqslant l=\begin{cases}\varepsilon, & q\leqslant\frac{1}{2}\\ \bar{\varepsilon}, & q>\frac{1}{2}\end{cases} \tag{2.37}$$

这里的精度控制数$l$指的是绝对误差限,亦可按相对误差限来控制

$$\left|\frac{x_{n+1}-\alpha}{x_{n+1}}\right|\leqslant\left|\frac{x_{n+1}-x_n}{x_{n+1}}\right|\leqslant\delta\text{(相对误差限)} \tag{2.38}$$

有的问题中还采用两种误差限并存控制如下:

当$|x_{n+1}|<c$且$|x_{n+1}-x_n|\leqslant l$时,终止迭代;否则继续迭代。

当$|x_{n+1}|\geqslant c$且$\left|\dfrac{x_{n+1}-x_n}{x_{n+1}}\right|\leqslant\delta$时，终止迭代；否则继续迭代。这里$c$为绝对误差限与相对误差限分界的控制数。

式(2.37)仅当$q<1$时才成立，当$q>1$的情况出现时，有可能出现假收敛的现象，这时虽然相邻两次近似值之差已满足(2.37)式，但$|x_{n+1}-\alpha|\leqslant l$并不满足，如图2.8所示。

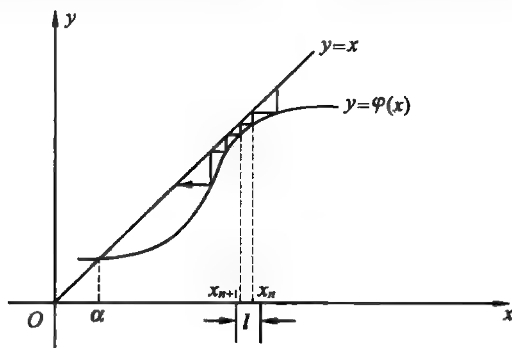



图 2.8

上界公式 2
$$|x_n-\alpha|\leqslant\frac{q^n}{1-q}|x_1-x_0| \tag{2.39}$$

**证**　式(2.39)可利用式(2.32)导出如下

$$\begin{aligned}|x_n-\alpha|&\leqslant\frac{q}{1-q}|x_n-x_{n-1}|=\frac{q}{1-q}|\varphi(x_{n-1})-\varphi(x_{n-2})|\\&=\frac{q}{1-q}|\varphi'(\xi_1)(x_{n-1}-x_{n-2})|\\&\leqslant\frac{q^2}{1-q}|x_{n-1}-x_{n-2}|\\&\leqslant\frac{q^3}{1-q}|x_{n-2}-x_{n-3}|\\&\cdots\\&\leqslant\frac{q^n}{1-q}|x_1-x_0|\end{aligned}$$
（证毕）

当
$$\frac{q^n}{1-q}|x_1-x_0|\leqslant\varepsilon \tag{2.40}$$
满足时，则$|x_n-\alpha|\leqslant\varepsilon$必成立。

式(2.40)亦可用来预估迭代次数：
$$n\ln q<\ln\frac{\varepsilon(1-q)}{|x_1-x_0|}$$

因为$\ln q<0$，所以有
$$n>\frac{\ln[\varepsilon(1-q)/|x_1-x_0|]}{\ln q} \tag{2.41}$$

选取满足上式的最小正整数便可作为迭代终止的计数控制参数。

**上界公式 3** $|x_n-\alpha|\leqslant\frac{|f(x_n)|}{m},\quad m\leqslant|f'(x)|,x\in[a,b]$ (2.42)

**证** 因 $f(\alpha)=0$，则有

$$f(x_n)-f(\alpha)=f(x_n)$$
$$f'(\eta)(x_n-\alpha)=f(x_n)$$
$$|x_n-\alpha|=\left|\frac{f(x_n)}{f'(\eta)}\right|\leqslant\frac{|f(x_n)|}{m},\quad m\leqslant|f'(x)|,x\in[a,b]$$ （证毕）

当

$$|f(x_n)|\leqslant m\varepsilon \tag{2.43}$$

满足时，就有 $|x_n-\alpha|\leqslant\varepsilon$ 成立。

有时亦有采用下面的判据

$$|f(x_n)|\leqslant\varepsilon \tag{2.44}$$

作为迭代终止的条件，这种判据是否合理？今分以下三种情况来分析它。

① 当 $|f'(x)|\approx1$ 时，式(2.44)与式(2.43)的效果相同，当式(2.44)满足时，就有 $|x_n-\alpha|\leqslant\varepsilon$ 成立。

② 当 $|f'(x)|\ll1$ 时，由 $|f(x_n)|\leqslant\varepsilon$ 可推得下式

$$|x_n-\alpha|\leqslant\frac{|f(x_n)|}{m}\leqslant\frac{\varepsilon}{m}=\varepsilon_1 \tag{2.45}$$

成立。由于 $\varepsilon<\varepsilon_1$，所以由式(2.45)知，迭代终止时的 $x_n$ 值的误差不一定能保证小于 $\varepsilon$。

③ 当 $|f'(x)|\gg1$ 时，由 $|f(x_n)|\leqslant\varepsilon$ 可推得下式

$$|x_n-\alpha|\leqslant\frac{|f(x_n)|}{m}=\frac{\varepsilon}{m}=\varepsilon_2 \tag{2.46}$$

成立。由于 $\varepsilon_2<\varepsilon$，所以由式(2.46)知，迭代终止时的 $x_n$ 值的误差比实际要求的误差 $\varepsilon$ 还要小，因而产生了不必要的过多次的迭代计算。综上可见，采用式(2.44)作为迭代终止的判据一般不是一种好方法。

**例 2.4** 对下列方程

$$x-\sin x-0.25=0 \tag{2.47}$$

使用迭代法，取三位小数计算其根的近似值，并估计其误差。

**解** 将式(2.47)分解为

$$x=\sin x+0.25 \tag{2.48}$$

由作图法可粗略知 $\alpha\in[0.9,1.5]$（图 2.9）。

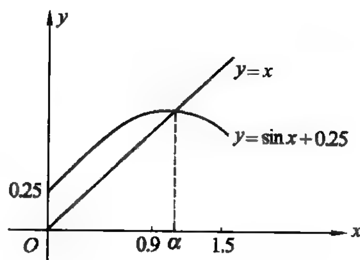



图 2.9

当 $x\in[0.9,1.5]$时，有

$|\varphi'(x)|=\cos x\leqslant\cos0.9=0.62<1$

所以迭代过程

$$x_{n+1}=\sin x_n+0.25 \tag{2.49}$$

必收敛。今取 $x_0=1.2$，按式(2.49)计算得 $x_4$

$$x_1=\sin1.2+0.250=0.932+0.250=1.182$$
$$x_2=\sin1.182+0.250=0.925+0.250=1.175$$
$$x_3=\sin1.175+0.250=0.923+0.250=1.173$$

$$x_4=\sin 1.173+0.250=0.922+0.250=1.172$$

则有$|x_4-x_3|=0.001$。按公式(2.32)估计得

$$|x_4-\alpha|\leqslant\frac{0.62}{1-0.62}\times 0.001=0.0016$$

计算 $x_4$ 的舍入误差限为 $2\times(0.5\times 10^{-3})=0.001$，最后得近似值 $x_4$ 的总误差为 $0.0016+0.001=0.0026<0.5\times 10^{-2}$，所以应取 $\alpha=1.17$。

注：如果把 $x=\varphi(x)$视为数学模型，则 $\alpha=\varphi(\alpha)$就是数学模型的精确解。而迭代公式 $x_{n+1}=\varphi(x_n)$应视为计算模型，它的精确解与数学模型精确解间的方法误差可应用前述的上界公式来估计。由式(2.39)可见，当迭代过程收敛时($q<1$)，随着 $n$ 的增加，$|x_n-\alpha|$的数值越来越小，最后 $x_n$ 可无限地趋近 $\alpha$，所以迭代法亦称为逐次逼近法。虽然迭代序列可逐次逼近解，由于受字长的限制，舍入误差必然存在，因此计算结果不可能达到任意的精度，在数字计算机上，最多只能达到机器精度。显见，只有当迭代法的数值解的舍入误差不大于迭代法的方法误差时，采用上述上界公式实现迭代终止控制才是可用的；且对于控制量小于或等于机器精度时都可能造成迭代死循环的情况，这时$|x_n-\alpha|<\varepsilon$ 的控制作用就会失效。

需要指出，每个迭代值的舍入误差甚至计算有误的误差在收敛的迭代过程($q<1$)中不会不断地累积到最终迭代值上，这是因为包含这类误差(设为$\in_n$)的迭代值 $x_n$ 可视为一个新的初值$x_n+\in_n$，经迭代一次得

$$|x_{n+1}-\alpha|\leqslant q|(x_n+\in_n)-\alpha|\leqslant q|x_n-\alpha|+q|\in_n|$$

可见误差$\in_n$在以后的迭代中会逐次地减小以致消失，从而导致最后一次迭代计算的舍入误差在舍入误差累积值中占主部地位的情况出现。因此，最终迭代值的总误差应由迭代公式的误差上界 $\varepsilon$ 和最后一次迭代计算的舍入误差之和组成，但$\in_n$ 的存在必然导致迭代次数的增加。

# §3　迭代公式的改进

为了使迭代过程收敛或在收敛较慢的情况下提高收敛速度，可设法提高初值的精度以减少迭代的次数。一般可使用简单的方法将初值提高到一定的精度；或者采用逼近公式来计算初值，具体方法可参看插值法和函数逼近两章内容。另一类方法就是选取适当的 $w$ 值以获得较佳的修改量 $wR_n$，从而达到快速收敛的目的。目前尚缺乏确定出这种 $w$ 值的直接方法，但却可以通过构建具有较小 $q$ 值的迭代函数的方法来确定出它的数值，下面介绍一些具体办法。

## 3.1　改变方程式法之一

本法是由分解式 $x=\varphi(x)$出发使用与 $\varphi(x)$有关的信息来构建具有更快收敛性的迭代函数的一类方法。

### 3.1.1　引入可选参数法

我们从 $x=\varphi(x)$出发，两边同时减去 $\theta x$，得到一个与 $f(x)=0$ 等价的方程

$$x-\theta x=\varphi(x)-\theta x$$

当 $\theta\neq 0$ 和 $\theta\neq 1$ 时，上式化为

$$x=\frac{1}{1-\theta}[\varphi(x)-\theta x]=\psi(x) \tag{2.50}$$

式中 $\theta$ 为可选的参数，它的取值应使 $|\psi'(x)|$ 尽可能地小为最佳。因

$$\psi'(x)=\frac{1}{1-\theta}[\varphi'(x)-\theta] \tag{2.51}$$

显见取 $\theta=\varphi'(x)$，$x\approx\alpha$ 比较合适。据选定的 $\theta$ 值便可建立如下的迭代公式

$$x_{n+1}=\frac{1}{1-\theta}[\varphi(x_n)-\theta x_n] \quad (n=0,1,2,\cdots) \tag{2.52}$$

在本法中，取

$$\begin{cases}\Delta x_n=\varphi(x_n)-x_n=\frac{1}{1-\theta}[\varphi(x_n)-\theta x_n]-x_n=\frac{1}{1-\theta}[\varphi(x_n)-x_n] \\ w=\frac{1}{1-\theta}\end{cases}$$

**例 2.5** 按式(2.52)解 $x=e^{-x}$。

**解**：因 $\alpha\in[0.5,0.6]$，所以有

$$-0.61=-e^{-0.5}\leqslant\varphi'(x)=-e^{-x}\leqslant-e^{-0.6}=-0.55$$

今粗取 $\theta=-0.6$，按(2.52)式建立迭代公式得

$$x_{n+1}=\frac{1}{1-(-0.6)}[e^{-x_n}-(-0.6)x_n]=\frac{1}{1.6}[e^{-x_n}+0.6x_n]$$

仍取 $x_0=0.5$，逐次计算得

$$x_1=0.566\,58, \quad x_2=0.567\,13, \quad x_3=0.567\,14$$

### 3.1.2 埃特肯法

(1) *方法描述*

为了获得较好的 $\theta$ 值，Aitken(1931 年)和 Steffensen(1933 年)分别设计出求取 $\theta$ 值的相同方法，现叙述如下。

首先将方程 $f(x)=0$ 分解为 $x=\varphi(x)$，然后由 $x_0$ 出发，迭代二次得到三个相邻的迭代值：$x_0$，$y_1=\varphi(x_0)$，$z_1=\varphi(y_1)$。取以下平均变化率作为 $\theta_0$

$$\theta_0=\frac{\varphi(y_1)-\varphi(x_0)}{y_1-x_0}=\frac{z_1-y_1}{y_1-x_0} \tag{2.53}$$

以 $x_0$ 和 $\theta=\theta_0$ 代入式(2.52)得

$$x_1=\frac{1}{1-\frac{z_1-y_1}{y_1-x_0}}\left[\varphi(x_0)-\frac{z_1-y_1}{y_1-x_0}\cdot x_0\right]=\frac{x_0z_1-y_1^2}{x_0-2y_1+z_1} \tag{2.54}$$

在求得 $x_1$ 的基础上，继续求取三个相邻迭代值：$x_1$，$y_2=\varphi(x_1)$，$z_2=\varphi(y_2)$，建立

$$\theta_1=\frac{\varphi(y_2)-\varphi(x_1)}{y_2-x_1}=\frac{z_2-y_2}{y_2-x_1}$$

同法以 $x_1$ 和 $\theta=\theta_1$ 代入式(2.52)得

$$x_2=\frac{x_1z_2-y_2^2}{x_1-2y_2+z_2} \tag{2.55}$$

仿此推导，直到 $|x_{n+1}-x_n|$ 之差满足精度要求为止。其一般公式可归纳为

$$x_{n+1}=\frac{x_nz_{n+1}-y_{n+1}^2}{x_n-2y_{n+1}+z_{n+1}}=\psi(x_n) \quad (n=0,1,2,\cdots) \tag{2.56}$$

其中

$$y_{n+1}=\varphi(x_n)$$
$$z_{n+1}=\varphi(y_{n+1})$$
$$\psi(x)=\frac{x\varphi[\varphi(x)]-[\varphi(x)]^2}{x-2\varphi(x)+\varphi[\varphi(x)]} \tag{2.57}$$

这个方法叫做埃特肯加速收敛法。Willers(1948 年)给出了它的几何解释(见图 2.10)。按照本法的计算过程，由 $x_0$ 出发，通过一次迭代得 $y_1=\varphi(x_0)$，这样便在曲线 $\varphi(x)$ 上得到点 $A(x_0,y_1)$。再迭代一次得 $z_1=\varphi(y_1)$，又可在曲线 $\varphi(x)$ 上得到点 $B(y_1,z_1)$。联结 $A$、$B$ 两点得到直线 $\overline{AB}$，设此直线与 $y=x$ 直线的交点为 $C(x_1,x_1)$，则 $C$ 点满足下式

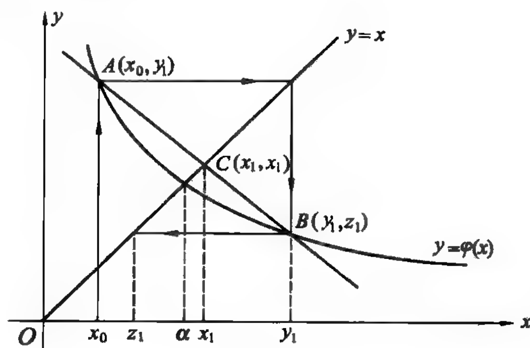



图 2.10

$$\frac{x_1-y_1}{x_1-x_0}=\frac{z_1-y_1}{y_1-x_0} \tag{2.58}$$

解出 $x_1$ 得

$$x_1=\frac{x_0z_1-y_1^2}{x_0-2y_1+z_1} \tag{2.59}$$

它与式(2.54)完全一致。可见埃特肯法就是通过建立直线 $\overline{AB}$ 代替曲线 $\varphi(x)$ 求取与 $y=x$ 交点的迭代方法。

在本法中，取

$$\begin{cases}\Delta x_n=\dfrac{1}{1-\theta_n}[\varphi(x_n)-x_n]\\ w_n=\dfrac{1}{1-\theta_n}\\ \theta_n=\dfrac{z_{n+1}-y_{n+1}}{y_{n+1}-x_n}\quad (y_{n+1}=\varphi(x_n),z_{n+1}=\varphi(y_{n+1}))\end{cases}$$

**例 2.6**　用埃特肯法解 $x=\mathrm{e}^{-x}$。

**解**　取 $x_0=0.5$，按式(2.56)计算得

$$\begin{cases}x_0=0.5\\ y_1=\mathrm{e}^{-0.5}=0.606\ 53\\ z_1=\mathrm{e}^{-0.606\ 53}=0.545\ 24\end{cases}$$

$$\begin{cases}x_1=\dfrac{0.5\times 0.545\ 24-0.606\ 53^2}{0.5-2\times 0.606\ 53+0.545\ 24}=0.567\ 62\\ y_2=\mathrm{e}^{-0.567\ 62}=0.566\ 87\\ z_2=\mathrm{e}^{-0.566\ 87}=0.567\ 30\end{cases}$$

$$\begin{cases}x_2=\dfrac{0.567\ 62\times 0.567\ 30-0.566\ 87^2}{0.567\ 62-2\times 0.566\ 87+0.567\ 30}=0.567\ 14\\ y_3=\mathrm{e}^{-0.567\ 14}=0.567\ 14\\ z_3=\mathrm{e}^{-0.567\ 14}=0.567\ 14\end{cases}$$

埃特肯法具有较高的收敛速度，其收敛的阶为 $r=2$。在有些情况下，使用简单迭代法发散而使用埃特肯法仍能获得收敛的效果。在使用式(2.56)计算时，要注意分母中相近数相减的问题。

**例 2.7** 求方程

$$x^5-x-0.2=0$$

在 $x_0=1$ 附近的根。

**解** 若按以下迭代公式

$$x_{n+1}=x_n^5-0.2=\varphi(x_n)$$

因 $|\varphi'(x_0)|=|5x^4|_{x_0=1}>1$，迭代过程发散。今采用埃特肯法解算得收敛的计算结果如表 2.2 所示。

**表 2.2**

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $x_n$ | 1 | 1.084 69 | 1.061 37 | 1.048 12 | 1.044 91 | 1.044 76 | 1.044 76 |
| $y_{n+1}$ | 0.8 | 1.301 51 | 1.146 90 | 1.064 90 | 1.045 65 | 1.044 75 | |
| $z_{n+1}$ | 0.127 68 | 3.534 54 | 1.784 39 | 1.169 44 | 1.050 06 | 1.044 69 | |

(2) 埃特肯法的收敛阶数

为了推证埃特肯法的收敛阶数，不失一般性，可令 $\alpha=0$。如果 $\alpha\neq 0$，可引入变换 $t=x-\alpha$，代入 $x=\varphi(x)$ 后得

$$t=\varphi(t+\alpha)-\alpha=\varphi^*(t) \tag{2.60}$$

则式(2.60)中的根归化为 $t=0$ 的情况。以下设 $x=\varphi(x)$ 的根为 $\alpha=0$，将 $\varphi(x)$ 在 $x=0$ 点展开得

$$\varphi(x)=\varphi(0)+\varphi'(0)x+\frac{\varphi''(\xi)}{2!}x^2,\qquad \xi\in(0,x)$$

因 $\varphi(0)=0$，上式化为

$$\varphi(x)=\varphi'(0)x+\frac{\varphi''(\xi)}{2!}x^2=cx+o(x) \tag{2.61}$$

其中 $c=\varphi'(0)$，$o(x)$ 为 $x$ 的高阶无穷小。运用上式继续推导得

$$\begin{aligned}\varphi[\varphi(x)]&=c[cx+o(x)]+o[cx+o(x)]\\&=c^2x+c\cdot o(x)+o(x)\\&=c^2x+o(x)+o(x)\\&=c^2x+o(x)\end{aligned} \tag{2.62}$$

$$\begin{aligned}[\varphi(x)]^2&=[cx+o(x)]^2=c^2x^2+2cxo(x)+[o(x)]^2\\&=c^2x^2+o(x^2)\end{aligned} \tag{2.63}$$

于是有

$$\begin{aligned}x\varphi[\varphi(x)]-[\varphi(x)]^2&=x[c^2x+o(x)]-[c^2x^2+o(x^2)]\\&=c^2x^2+xo(x)-c^2x^2-o(x^2)=o(x^2)\end{aligned} \tag{2.64}$$

$$\begin{aligned}x-2\varphi(x)+\varphi[\varphi(x)]&=x-2[cx+o(x)]+[c^2x+o(x)]\\&=x-2cx-2o(x)+c^2x+o(x)\\&=(c-1)^2x+o(x)\end{aligned} \tag{2.65}$$

将式(2.64)与式(2.65)代入式(2.57)中得

$$\begin{cases}\psi(x)=\dfrac{o(x^2)}{(c-1)^2x+o(x)}=o(x)=kx^2 \quad (k\text{ 为常数})\\ \psi'(x)=2kx \\ \psi''(x)=2k\end{cases} \tag{2.66}$$

因此推知

$$\psi(o)=0,\ \psi'(o)=0,\ \psi''(o)\neq 0$$

证得埃特肯法收敛的阶为 $r=2$。

### 3.1.3　组合法

当取定 $R_n=\varphi(x_n)-x_n$ 时，式(2.9)化为

$$\begin{aligned}x_{n+1}&=x_n+w[\varphi(x_n)-x_n]=(1-w)x_n+w\varphi(x_n)\\&=(1-w)x_n+wy_{n+1}\end{aligned} \tag{2.67}$$

其中 $y_{n+1}=\varphi(x_n)$。它实际上就是相邻两个迭代值 $x_n$，$y_{n+1}$ 按 $(1-w)$ 与 $w$ 之比的组合公式。如取 $w=1/2$，就得到以下迭代公式

$$\begin{cases}x_{n+1}=\dfrac{x_n+y_{n+1}}{2}\\ y_{n+1}=\varphi(x_n)\end{cases} \tag{2.68}$$

它取相邻两个迭代值的算术平均值作为新的近似值(图 2.11)。当迭代序列中的各次近似值在根的两侧往复地趋近 $\alpha$ 时，使用式(2.68)除能加快收敛外，还可有效地防止死循环的出现。在本法中，取

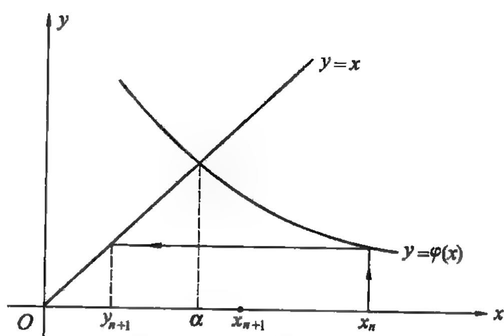


图 2.11

$$\begin{cases}\Delta x_n=\dfrac{x_n+\varphi(x_n)}{2}-x_n=\dfrac{1}{2}[\varphi(x_n)-x_n]\\ w=\dfrac{1}{2}\end{cases}$$

**例 2.8**　用式(2.68)解算 $x=e^{-x}$。

**解**　取初值 $x_0=0.5$，按式(2.68)计算得表 2.3。

**表 2.3**

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $x_n$ | 0.5 | 0.553 27 | 0.564 17 | 0.566 50 | 0.567 00 | 0.567 11 | 0.567 14 |
| $y_{n+1}$ | 0.606 53 | 0.575 07 | 0.568 83 | 0.567 51 | 0.567 22 | 0.567 16 | 0.567 14 |

本法可视为两个迭代值的组合法，而埃特肯法则可视为三个迭代值的组合法，组合公式就是一种化粗为精的公式，常可达到优劣数据互补、取长补短的效果。

## 3.2　改变方程式法之二

残差 $R_n$ 亦可由方程 $f(x)=0$ 出发来获得，令 $R'_n=0-f(x_n)$，则有关系式

$$R_n=\lambda'R'_n=\lambda'[0-f(x_n)]=-\lambda'f(x_n)$$

相应的迭代公式为

$$x_{n+1}=x_n-\lambda' f(x_n)\qquad (\lambda'\text{为常数}) \tag{2.69}$$

与式(2.6)对应的迭代公式为

$$x_{n+1}=x_n-w\lambda' f(x_n)=x_n-\lambda f(x_n)=\psi(x_n) \tag{2.70}$$

式中，$\lambda=w\lambda'$为常数，式(2.70)可视为方程 $f(x)=0$ 作 $x=\psi(x)$分解后所构成的简单迭代法。从式(2.70)可见，$\lambda$ 的选取与 $f(x)$的信息有关，在本法中，取

$$\begin{cases} x=x-\lambda f(x)=\psi(x) \\ \Delta x_n=-\lambda f(x_n) \end{cases} \tag{2.71}$$

今选择 $\lambda$ 值，使$|\psi'(x)|<1$，以达到收敛的目的。因

$$|\psi'(x)|=|1-\lambda f'(x)| \tag{2.72}$$

由上式可见，$\lambda$ 值的选取与 $f'(x)$的大小有关，为此，先对 $f'(x)$在$[a,b]$内的值作出以下的估计

$$0<m\leqslant f'(x)\leqslant M$$

这里假设 $f'(x)>0$，若 $f'(x)<0$ 时，可对原方程乘以负号就可化成上述情况。

为使$|\psi'(x)|<1$，就必须要求$|1-\lambda m|<1$ 与$|1-\lambda M|<1$ 同时满足。为使$|1-\lambda m|<1$满足，就必须要求 $1-\lambda m<1$ 与$-(1-\lambda m)<1$ 同时满足，由此获得

$$0<\lambda<\frac{2}{m} \tag{2.73}$$

同法，由$|1-\lambda M|<1$ 获得解

$$0<\lambda<\frac{2}{M} \tag{2.74}$$

则式(2.74)为公共解，亦即所选 $\lambda$ 值应满足式(2.74)。由式(2.72)可见，若取

$$\lambda=\frac{1}{f'(x)},\qquad x\in[a,b] \tag{2.75}$$

可获得较小的$|\psi'(x)|$值，通常可取为

$$f'(x)=\frac{m+M}{2}$$

$$\lambda^*=\frac{1}{\frac{m+M}{2}}=\frac{2}{m+M} \tag{2.76}$$

显然 $0<\lambda^*<\frac{2}{M}$，相应的迭代公式为

$$x_{n+1}=x_n-\frac{2}{m+M}f(x_n) \tag{2.77}$$

**例 2.9** 按式(2.77)解 $x=\mathrm{e}^{-x}$。

**解** 在本例中，$x\in[0.5,0.6]$，有

$$f(x)=x-\mathrm{e}^{-x}$$

$$f'(x)=1+\mathrm{e}^{-x}$$

$$m=f'(0.6)=1+\mathrm{e}^{-0.6}=1.55$$

$$M=f'(0.5)=1+\mathrm{e}^{-0.5}=1.61$$

$$\lambda^*=\frac{2}{1.55+1.61}=0.63$$

代入式(2.77)得

$$x_{n+1}=x_n-0.63(x_n-e^{-x_n})=0.37x_n+0.63e^{-x_n} \tag{2.78}$$

取 $x_0=0.5$,按式(2.78)计算得

$$x_0=0.5, x_1=0.567\ 11, x_2=0.567\ 14, x_3=0.567\ 14$$

如果我们把式(2.71)中的迭代函数改写为

$$x=x-\lambda f(x)=x-\lambda(x-\varphi(x))=(1-\lambda)x+\lambda\varphi(x) \tag{2.79}$$

并与式(2.50)

$$x=\frac{-\theta}{1-\theta}x+\frac{1}{1-\theta}\varphi(x) \tag{2.80}$$

比较,就知 $\lambda$ 与 $\theta$ 间有以下关系

$$\lambda=\frac{1}{1-\theta}$$

$$\theta=1-\frac{1}{\lambda} \tag{2.81}$$

可见式(2.50)与式(2.71)间可以相互转换。

## 3.3　牛顿迭代法

### 3.3.1　方法的描述

为使迭代公式

$$x_{n+1}=x_n-\lambda f(x_n) \qquad (n=0,1,2,\cdots)$$

在每一次迭代中均有较高的收敛速度,$\lambda$ 不再取固定值而改为变值如下

$$\lambda=\frac{1}{f'(x_n)} \tag{2.82}$$

于是得到以下迭代公式

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)} \qquad (n=0,1,2,\cdots) \tag{2.83}$$

这就是牛顿迭代法。它具有明显的几何意义(图 2.12),我们建立 $(x_n, f(x_n))$ 点的 $y=f(x)$ 的切线方程

$$y=f(x_n)+f'(x_n)(x-x_n) \tag{2.84}$$

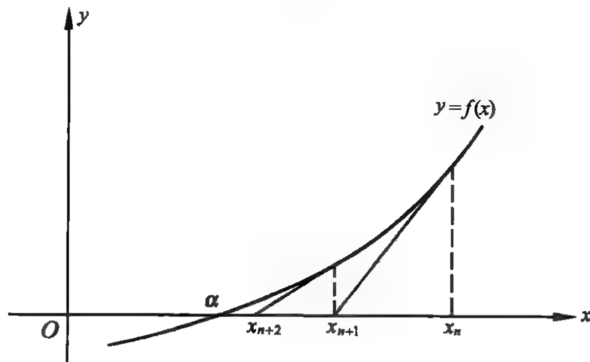



图 2.12

设其与 $x$ 轴的交点为$(x_{n+1},0)$，代入式(2.84)得

$$0=f(x_n)+f'(x_n)(x_{n+1}-x_n)$$

解得

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)} \tag{2.85}$$

它与牛顿迭代法完全一致，因此牛顿迭代法也称为切线法。在本法中，取

$$\begin{cases}\Delta x_n=\dfrac{f(x_n)}{f'(x_n)}\\ \lambda=\dfrac{1}{f'(x_n)}\end{cases}$$

**例 2.10** 按牛顿迭代法解 $x=e^{-x}$。

**解** 因

$$f(x)=xe^x-1$$
$$f'(x)=e^x(1+x)$$

按式(2.83)得如下迭代公式

$$x_{n+1}=x_n-\frac{x_ne^{x_n}-1}{e^{x_n}(1+x_n)}=x_n-\frac{x_n-e^{-x_n}}{1+x_n}$$

取 $x_0=0.5$，逐次计算得

$$x_0=0.5,x_1=0.571\,02,x_2=0.567\,16,x_3=0.567\,14,x_4=0.567\,14$$

在切线法中，若出现 $f'(x_n)=0$ 时，即切线与 $x$ 轴无交点的情况下，方法就失效了。在这种情况下，可处理如下。

设在切线法中，出现 $f'(x_n)=0$ 且有 $f(x_n)\cdot f''(x_n)<0$ 成立，显然方程有根 $\xi$ 与 $\xi'$ 存在(图 2.13)。今将方程 $f(x)=0$ 中的 $f(x)$ 按台劳公式展开并截取其前三项后得

$$f(x_n)+\frac{1}{2}f''(x_n)(x-x_n)^2=0 \tag{2.86}$$

解得

$$\begin{cases}\eta=x_n-\sqrt{-\dfrac{2f(x_n)}{f''(x_n)}}\\ \eta'=x_n+\sqrt{-\dfrac{2f(x_n)}{f''(x_n)}}\end{cases} \tag{2.87}$$

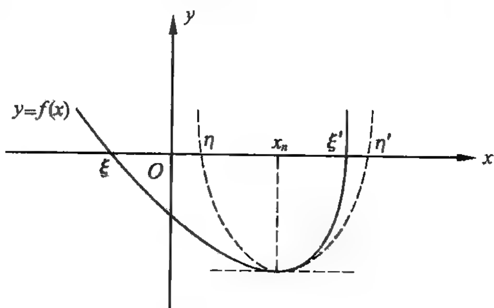


图 2.13

上述二根实际上就是抛物线

$$y=f(x_n)+\frac{1}{2}f''(x_n)(x-x_n)^2 \tag{2.88}$$

与 $x$ 轴交点的横坐标。以下可取 $\eta$ 或 $\eta'$ 为初值，使用切线法分别求取其邻近的根 $\xi$ 和 $\xi'$。

### 3.3.2 牛顿迭代法的收敛性

牛顿迭代法具有较高的收敛速度，它的收敛阶数为 $r=2$；在重根情况下，$r=1$(见3.6.1节)。其迭代函数为

$$\psi(x)=x-\frac{f(x)}{f'(x)} \tag{2.89}$$

于是有
$$\psi'(x)=1-\frac{[f'(x)]^2-f(x)f''(x)}{[f'(x)]^2}=\frac{f(x)f''(x)}{[f'(x)]^2} \tag{2.90}$$
当 $f'(x)\neq 0$ 及具有连续的 $f''(x)$ 时，只要 $x_0$ 充分地接近 $\alpha$，就可以使 $f(x_0)$ 足够小而使 $|\psi'(x)|<1$，保证了牛顿迭代法的收敛性。牛顿迭代法的局部收敛性较强，只有初值充分地接近 $\alpha$，才能确保迭代序列的收敛性。为了放宽对局部收敛性的限制，必须再增加条件建立以下收敛的充分条件。

**定理 2.5**　若 $f(x)$ 在 $[a,b]$ 上二阶导数存在，且满足

① $f(a)f(b)<0$；

② $f'(x)\neq 0$；

③ $f''(x)$ 不变号；

④ 初值 $x_0$ 满足 $f(x_0)f''(x_0)>0$，则牛顿迭代法收敛。　(2.91)

证明从略，只作它的几何意义说明如下：条件①保证了根的存在性。条件②表明函数单调变化，$[a,b]$ 内根唯一。③中 $f''(x)$ 不变号表示 $f(x)$ 的图形在 $[a,b]$ 上的凹向不变。条件③和④一起保证了每一次迭代值都界于 $[a,b]$ 内(见图 2.14)。

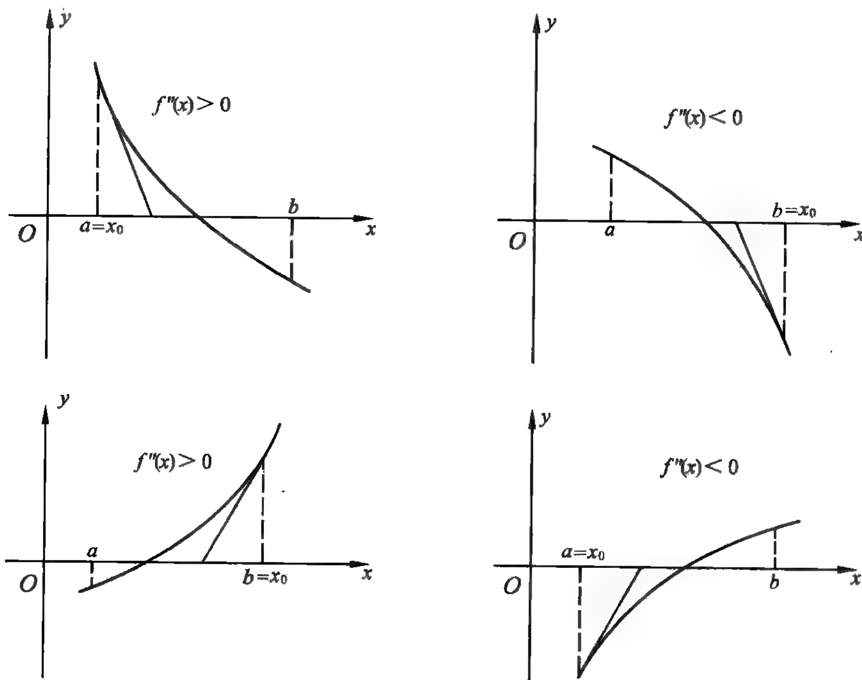



图 2.14

在不满足上述收敛充分条件时，有可能导致迭代值远离所求根的情况或死循环的情况(图 2.15)。

**例 2.11**　研制求取 $\sqrt{c}\,(0<c<1)$ 的快精算法。

在本例中，全部数值均采用绝对值小于 1 的定点数。为求取 $\sqrt{c}$，令 $x=\sqrt{c}$，则有
$$f(x)=x^2-c=0 \tag{2.92}$$

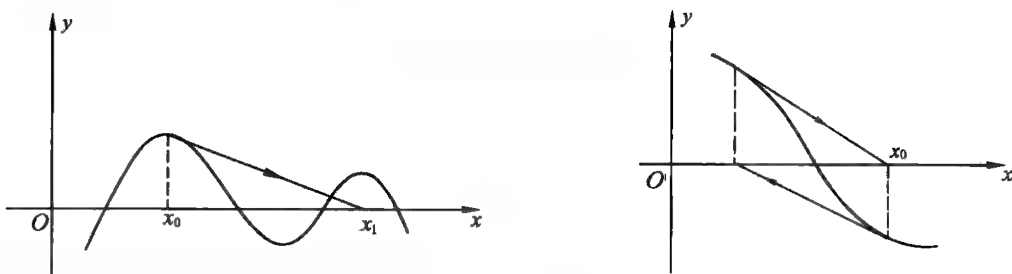



图 2.15

今采用牛顿迭代法解上述方程，得迭代公式

$$x_{n+1}=x_n-\frac{x_n^2-c}{2x_n}=\frac{1}{2}\left(x_n+\frac{c}{x_n}\right) \tag{2.93}$$

由式(2.90)知，若$|f''(x)|$愈小、$|f'(x)|$愈大、$x_0$ 愈接近于 $\alpha$，则$|\psi'(x)|$就愈小，收敛就愈快。在本例中，$f'(x)=2x, f''(x)=2$。当 $c$ 较小时，$x$ 也较小，则 $f'(x)$也较小，这时收敛较慢。另外，由于字长有限，舍入误差在计算结果中占有较大的比重而影响结果的精度。以上两个弊病可用扩大 $c$ 值的办法来解决，具体方法就是将 $c$ 值(非规格化数)左移成规格化数。在采用二进制数的计算机上，左移一次就是将 $c$ 扩大 2 倍，设 $c$ 左移 $N$ 次后成为规格化数 $m$，则有以下关系式

$$2^N c=m \qquad \left(\frac{1}{2}\leqslant m<1\right)$$

$$c=\frac{m}{2^N}$$

$$x=\sqrt{c}=\begin{cases}\dfrac{\sqrt{m}}{2^{\frac{N}{2}}}, & N\text{ 为偶数}\\[2ex] \dfrac{\sqrt{m}}{2^{\frac{N-1}{2}}\cdot 2^{\frac{1}{2}}}=\dfrac{\sqrt{m}/\sqrt{2}}{2^{\frac{N-1}{2}}}, & N\text{ 为奇数}\end{cases} \tag{2.94}$$

这样处理后，求取$\sqrt{c}$的问题转化为求取$\sqrt{m}$的问题。对于$\sqrt{m}$，可令 $f(y)=y^2-m=0$ 后继续采用牛顿迭代法求解

$$y_{n+1}=\frac{1}{2}\left(y_n+\frac{m}{y_n}\right) \qquad (n=0,1,2,\cdots) \tag{2.95}$$

在求得满足精度要求的$\sqrt{m}$值后，根据式(2.94)，应对它作 $2^{\frac{N}{2}}$ 或 $2^{\frac{N-1}{2}}$ 的除法运算。在计算机上，除以 2 可对数值右移一次来实现，这里的右移次数为$\frac{N}{2}$或$\frac{N-1}{2}$次。因 $N$ 为规格化 $c$ 时的左移次数，可用一个计数单元从其最末位逐次累加获得。对 $N$ 或 $N-1$ 除以 2 的运算只需将 $N$ 的数值右移一位即得。在右移中，若 $N$ 的末位数字为 1 时就自动被移掉，不必在右移前作 $N-1$ 的运算。

剩下要解决的问题是初值 $y_0$ 的确定问题。由式(2.91)知，$y_0$ 应满足 $f(y_0)f''(y_0)>0$ 的要求，即

$$2(y_0^2-m)>0$$

$$y_0>\sqrt{m} \tag{2.96}$$

为了用最简便的方法获取较好的初值,我们以间距 $h$ 将区间 $[\frac{1}{2},1]$ 等分为 $n$ 个子区间

$$[m_0,m_1],[m_1,m_2],\cdots,[m_{n-1},m_n] \tag{2.97}$$

式中,$m_0=\frac{1}{2}$,$m_n=1$。各分点的数值为

$$m_i=\frac{1}{2}+ih \qquad (i=0,1,2,\cdots,n) \tag{2.98}$$

分点所对应的函数值为 $y_i=\sqrt{m_i}(i=0,1,2,\cdots,n)$,这些函数值可依次存储起来备用,其值可按下标 $i$ 来取定。

对于给定的 $m$ 值,设它位于某个子区间 $[m_i,m_{i+1}]$ 内,按式(2.96)的要求,可取定该子区间右端点的函数值为初值:

$$y_0=\sqrt{m_{i+1}} \qquad (>\sqrt{m}) \tag{2.99}$$

初值 $y_0$ 的精度取决于 $h$ 的大小,因此确定 $h$ 大小是个关键性问题。它要根据最终计算结果所要求的精度和预定的迭代次数来确定。迭代次数的多少又取决于对求解时间的实际需求,若迭代次数要求少,则达到最终结果的精度所要求的初值精度就高,相应的 $h$ 应取得小。在计算机上,为使选取初值的工作更加简便,选用如下形式的 $h$

$$h=2^{-k} \qquad (k\text{ 为}\geqslant 1\text{ 的正整数}) \tag{2.100}$$

是最为合宜的,这时 $m$ 位于 $[m_i,m_{i+1}]$ 区间的判断工作就可归结为如下求取 $i$ 的问题,其计算公式为

$$i=\left[\frac{m-m_0}{2^{-k}}\right]=\left[2^k\cdot\left(m-\frac{1}{2}\right)\right]=\left[2^k\bar{m}\right] \tag{2.101}$$

式中,$\bar{m}=m-\frac{1}{2}$,[ ]为取整值运算符。$\bar{m}$ 值只需将 $m$ 的小数后第一位数字 1 置为 0 即得。而 $i$ 值就是 $\bar{m}$ 左移 $k$ 位后的整值部分。如果保持 $\bar{m}$ 不动,则可截取 $\bar{m}$ 小数后的 $k$ 位数字来构成整值 $i$。根据 $i$ 值便可取定初值 $y_0=\sqrt{m_{i+1}}$。

由式(2.100)可见,$h$ 的大小取决于 $k$ 的大小,而 $k$ 的大小可根据结果的精度要求及对迭代次数的多少通过试算来确定。

### 3.3.3 切线法的变形使用

(1) 简化切线法

为避免频繁地计算导数值 $f'(x_n)$,可将它取为固定值,比如取为 $f'(x_0)$,这时得以下迭代公式

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_0)} \tag{2.102}$$

称为简化切线法或固定斜率的切线法,它的几何意义如图 2.16 所示。

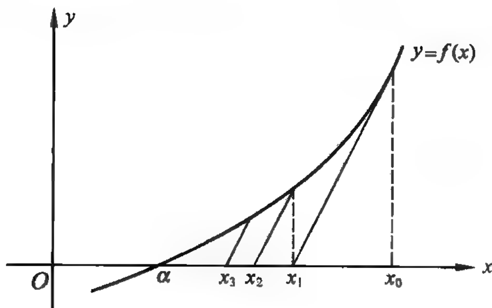



图 2.16

在本法中,取

$$\begin{cases}\Delta x_n=-\dfrac{f(x_n)}{f'(x_0)}\\ \lambda=\dfrac{1}{f'(x_0)}\end{cases}$$

更一般地，可取

$$f'(x_n)=c\ (\text{常数}) \tag{2.103}$$

则迭代公式成为

$$x_{n+1}=x_n-\frac{f(x_n)}{c} \tag{2.104}$$

称为推广的简化切线法。这时 $c$ 值应满足下式

$$|\psi'(x)|=\left|1-\frac{f'(x)}{c}\right|<1 \tag{2.105}$$

满足式(2.105)的解为

$$0<\frac{f'(x)}{c}<2 \tag{2.106}$$

可见当 $c$ 与 $f'(x)$ 同号且满足上述不等式时，推广的简化切线法是收敛的。

(2) 修正的切线法

此法进行如下，由 $x_n$ 近似值出发，按简化切线法迭代 $m$ 次后得 $x_{n+1}$，以 $x_{n+1}$ 代替 $x_n$，重复上述过程，直到满足精度要求为止。

(3) 牛顿下山法

由于牛顿迭代法的收敛性太依赖于初值 $x_0$，如果 $x_0$ 偏离 $\alpha$ 较远时，则可能发散。为了扩大收敛范围，可在牛顿迭代公式中引进参数 $\tilde{\lambda}$

$$x_{n+1}=x_n-\tilde{\lambda}\frac{f(x_n)}{f'(x_n)} \tag{2.107}$$

适当选择 $\lambda$ 的值，使满足下面的下山条件

$$|f(x_{n+1})|<|f(x_n)| \tag{2.108}$$

使之达到 $|f(x_0)|>|f(x_1)|>\cdots$ 单调下降的目的。称这种算法为牛顿下山法，其计算过程如下。

首先取 $\tilde{\lambda}=1$，在 $x_n$ 基础上用牛顿迭代法迭代一次得 $x_{n+1}$，然后检查下山条件式(2.108)是否满足？若下山条件满足，则继续用牛顿迭代法计算，当满足精度要求时停止迭代，获得最终结果。若下山条件不满足，用不断地修改 $\tilde{\lambda}$ 值的办法使下山条件得到满足为止，并取下山条件满足时的迭代值作为 $x_{n+1}$。以下计算又改为 $\tilde{\lambda}=1$ 时的牛顿迭代法，仿上进行，直至 $x_{n+1}$ 满足精度要求为止。

在上述计算过程中，用 $\lambda$ 来修改原切线的斜率，以改变 $|f(x_{n+1})|$ 与 $|f(x_n)|$ 的偏离程度，最后达到下山的目的。这里 $\tilde{\lambda}$ 称为下山因子，$\tilde{\lambda}$ 值太小时，相邻两次迭代值的差别不大，其调节效果不明显，因此可设置一个下限 $\varepsilon_1$，在 $[\varepsilon_1,1]$ 范围内进行选值，选值方式一般可依次地取

$$1,\frac{1}{2},\frac{1}{4},\cdots,\frac{1}{2^r}>\varepsilon_1 \tag{2.109}$$

如果在上述范围内仍然没有使下山条件满足的 $\tilde{\lambda}$ 值，则应另选初值，重新进行迭代。

为了简化上述 $\tilde{\lambda}$ 值的调试过程，可将公式(2.107)做以下变形

$$\begin{aligned}x_{n+1}&=x_n-\tilde{\lambda}\frac{f(x_n)}{f'(x_n)}+\tilde{\lambda}x_n-\tilde{\lambda}x_n\\&=(1-\tilde{\lambda})x_n+\tilde{\lambda}\left[x_n-\frac{f(x_n)}{f'(x_n)}\right]\end{aligned} \tag{2.110}$$

令 $y_{n+1}=x_n-\dfrac{f(x_n)}{f'(x_n)}$,代入上式后得

$$x_{n+1}=(1-\tilde{\lambda})x_n+\tilde{\lambda}y_{n+1} \tag{2.111}$$

它是一次牛顿迭代法中,旧、新迭代值 $x_n,y_{n+1}$ 按 $(1-\tilde{\lambda})$ 与 $\tilde{\lambda}$ 比值的组合公式,在保持 $x_n,y_{n+1}$ 不变的基础上,取

$$\tilde{\lambda}=\tilde{\lambda}_i=\frac{1}{2^i}\qquad(i=0,1,2,\cdots) \tag{2.112}$$

计算相应的 $x_{n+1}^{(i)}$

$$x_{n+1}^{(i)}=(1-\tilde{\lambda}_i)x_n+\tilde{\lambda}_i y_{n+1} \tag{2.113}$$

直到下山条件

$$|f(x_{n+1}^{(i)})|<|f(x_n)| \tag{2.114}$$

满足为止。在本法中,取

$$\begin{cases}\Delta x_n=-\tilde{\lambda}_i\dfrac{f(x_n)}{f'(x_n)}\\ \tilde{\lambda}=\dfrac{\tilde{\lambda}_i}{f'(x_n)}\end{cases}\quad(\tilde{\lambda}_i\text{ 据下山条件被满足来择定})$$

**例 2.12**　用牛顿下山法求方程的根

$$f(x)=x^3-x-1=0 \tag{2.115}$$

**解**　方程(2.115)在[0,1.5]内有根,如取 $x_0=0.6$,用牛顿迭代法计算得 $y_1=17.9$,显然迭代不收敛,且下山条件不满足:$|f(17.9)|>|f(x_0)|$。以下修改 $\tilde{\lambda}$ 值,使下山条件满足

取 $\tilde{\lambda}_1=\dfrac{1}{2}$, $x_1^{(1)}=(1-\dfrac{1}{2})x_0+\dfrac{1}{2}y_1=9.25$, $|f(9.25)|>|f(x_0)|$

取 $\tilde{\lambda}_2=\dfrac{1}{2^2}$, $x_1^{(2)}=(1-\dfrac{1}{2^2})x_0+\dfrac{1}{2^2}y_1=4.925$, $|f(4.925)|>|f(x_0)|$

取 $\tilde{\lambda}_3=\dfrac{1}{2^3}$, $x_1^{(3)}=(1-\dfrac{1}{2^3})x_0+\dfrac{1}{2^3}y_1=2.7625$, $|f(2.7625)|>|f(x_0)|$

取 $\tilde{\lambda}_4=\dfrac{1}{2^4}$, $x_1^{(4)}=(1-\dfrac{1}{2^4})x_0+\dfrac{1}{2^4}y_1=1.68125$, $|f(1.68125)|>|f(x_0)|$

取 $\tilde{\lambda}_5=\dfrac{1}{2^5}$, $x_1^{(5)}=(1-\dfrac{1}{2^5})x_0+\dfrac{1}{2^5}y_1=1.140625$, $|f(1.140625)|<|f(x_0)|$

这时下山条件已满足,取 $x_1=1.140625$。以下继续按牛顿迭代法迭代($\tilde{\lambda}=1$)

$$x_2=1.140625-\frac{f(1.140625)}{f'(1.140625)}=1.366814,\ |f(x_2)|<|f(x_1)|$$

$$x_3=1.366814-\frac{f(1.366814)}{f'(1.366814)}=1.32628,\ |f(x_3)|<|f(x_2)|$$

$$x_4=1.32628-\frac{f(1.32628)}{f'(1.32628)}=1.32472,\ |f(x_4)|<|f(x_3)|$$

$$x_5=1.32472-\frac{f(1.32472)}{f'(1.32472)}=1.32472$$

已达到五位小数一致的结果。

## 3.4 弦截法

### 3.4.1 单点弦截法

为避免牛顿迭代法中导数 $f'(x_n)$ 的计算，可用以下平均变化率

$$f'(x_n)\approx\frac{f(x_n)-f(x_0)}{x_n-x_0} \tag{2.116}$$

来代替 $f'(x_n)$，于是得到如下迭代公式

$$x_{n+1}=x_n-\frac{f(x_n)}{f(x_n)-f(x_0)}(x_n-x_0)=\frac{x_0f(x_n)-x_nf(x_0)}{f(x_n)-f(x_0)} \quad (n=1,2,\cdots) \tag{2.117}$$

称为单点弦截法。在本法中，取

$$\begin{cases}\Delta x_n=-\dfrac{f(x_n)}{f(x_n)-f(x_0)}(x_n-x_0)\\ \lambda=\dfrac{1}{\dfrac{f(x_n)-f(x_0)}{x_n-x_0}}\end{cases}$$

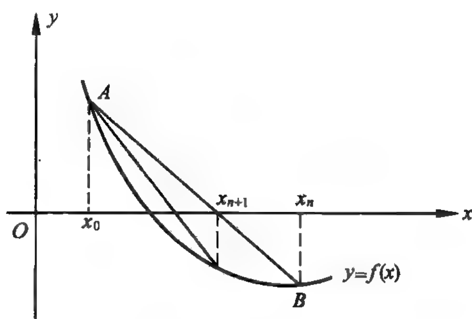



图 2.17

单点弦截法具有明显的几何意义(图 2.17)，它是用联结点 $A(x_0,y_0)$ 与点 $B(x_n,y_n)$ 的直线 $\overline{AB}$ 代替 $y=f(x)$ 曲线求取与横轴交点作为近似值 $x_{n+1}$ 的方法，以后再过 $(x_0,y_0)$ 与 $(x_{n+1},y_{n+1})$ 两点作直线求取与横轴的交点作为 $x_{n+2}$ 等等。其中 $(x_0,y_0)$ 是一个固定点，称为不动点，另一点则不断更换，故名单点弦截法。

单点弦截法的迭代函数为

$$\psi(x)=x-\frac{f(x)}{f(x)-f(x_0)}(x-x_0) \tag{2.118}$$

因 $(x_0,y_0)$ 为不动点，则可视迭代序列中的每个值为一个初始近似值，不妨假定 $x_n$ 为初值，来讨论 $|\psi'(x)|$ 的大小，按式(2.118)得

$$|\psi'(x_n)|=\left|1+\frac{f'(x_n)f(x_0)}{[f(x_n)-f(x_0)]^2}(x_n-x_0)-\frac{f(x_n)}{f(x_n)-f(x_0)}\right| \tag{2.119}$$

当 $x_n$ 充分地接近 $\alpha$ 时，有 $x_n\approx\alpha$，$f(x_n)\approx0$，上式可近似为

$$\begin{aligned}|\psi'(x_n)|&\approx\left|1+\frac{f'(\alpha)f(x_0)}{[f(x_0)]^2}(\alpha-x_0)\right|\\ &=\left|1+\frac{f'(\alpha)}{f(x_0)-f(\alpha)}(\alpha-x_0)\right|\\ &=\left|1+\frac{f'(\alpha)}{f'(\xi)(x_0-\alpha)}(\alpha-x_0)\right|\\ &=\left|1-\frac{f'(\alpha)}{f'(\xi)}\right| \qquad \xi\in(x_0,\alpha)\end{aligned} \tag{2.120}$$

由式(2.120)可见，只要 $x_n$ 充分地接近 $\alpha$，且 $f'(x)\neq0$ 及变化不大时，就有 $|\psi'(x_n)|<1$ 成立。可以证明(见第五章反插值法)，单点弦截法具有收敛的阶 $r=1$，即具有线性收敛速度。

与牛顿迭代法类似，为扩大局部收敛范围，有以下收敛的充分条件。

**定理 2.6** 若 $f(x)$ 在 $[a,b]$ 上二阶导数存在，且满足

① $f(a)f(b)<0$；

② $f'(x)\neq 0$；

③ $f''(x)$不变号；

④ 不动点 $x_0$ 满足 $f(x_0)f''(x_0)>0$，$x_1$ 与 $x_0$ 点的函数值相异，则单点弦截法收敛。

**例 2.13**　用单点弦截法解 $f(x)=xe^x-1=0$。

**解**　因 $\alpha\in[0.5,0.6]$及

$$f(x)=xe^x-1$$
$$f'(x)=e^x(1+x)$$
$$f''(x)=e^x(2+x)$$

计算

$$f(0.5)f''(0.5)<0,\quad f(0.6)f''(0.6)>0$$

所以取$[0.6,f(0.6)]$为不动点，得 $x_0=0.6$，另一端点取为 $x_1=0.5$。按式(2.117)计算得：

$$x_2=\frac{0.6\times f(0.5)-0.5\times f(0.6)}{f(0.5)-f(0.6)}=0.565\ 32$$

$$x_3=\frac{0.6\times f(x_2)-0.565\ 32\times f(0.6)}{f(x_2)-f(0.6)}=0.567\ 09$$

$$x_4=\frac{0.6\times f(x_3)-0.567\ 09\times f(0.6)}{f(x_3)-f(0.6)}=0.567\ 14$$

$$x_5=\frac{0.6\times f(x_4)-0.567\ 14\times f(0.6)}{f(x_4)-f(0.6)}=0.567\ 14$$

在单点弦截法中，不动点若不满足条件 $f(x_0)f''(x_0)>0$，则可能会出现各次迭代值在 $\alpha$ 附近左右摆动现象，如图 2.18 所示。

### 3.4.2　双点弦截法

若把单点弦截法中的不动点改为变动点$(x_{n-1},y_{n-1})$，则得到下面的双点弦截法的迭代公式

$$x_{n+1}=x_n-\frac{f(x_n)}{f(x_n)-f(x_{n-1})}(x_n-x_{n-1})$$
$$=\frac{x_{n-1}f(x_n)-x_nf(x_{n-1})}{f(x_n)-f(x_{n-1})}\qquad(n=1,2,\cdots)\tag{2.121}$$

其几何意义如图 2.19 所示。

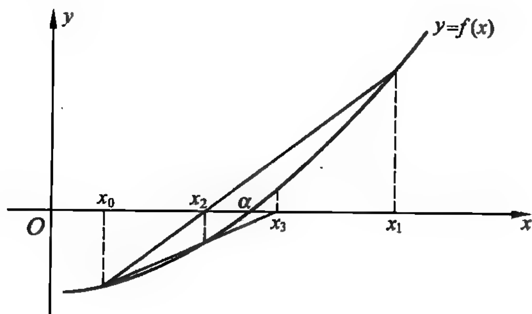



图 2.18

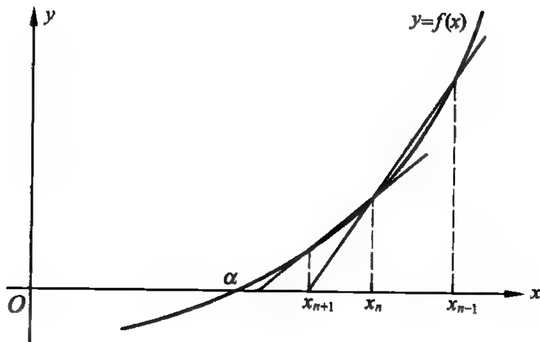



图 2.19

在本法中，取

$$\begin{cases}\Delta x_n = -\dfrac{f(x_n)}{f(x_n)-f(x_{n-1})}(x_n - x_{n-1}) \\ \lambda = \dfrac{1}{\dfrac{f(x_n)-f(x_{n-1})}{x_n - x_{n-1}}}\end{cases}$$

与单点弦截法类似，双点弦截法有以下收敛的充分条件。

**定理 2.7** 若 $f(x)$在$[a,b]$上有直至二阶的连续导数，且满足

① $f(a)f(b)<0$；

② $f'(x)\neq 0$。

则对任意 $x_0, x_1 \in [a,b]$双点弦截法均收敛。

可以证明，双点弦截法在收敛时，其收敛的阶为 $r=\frac{1}{2}(1+\sqrt{5})\approx 1.618$(见第五章反插值法)。在双点弦截法使用中，须注意到分母中相近数相减的问题及有可能当 $n$ 增加时出现的近似值在 $\alpha$ 左右明显地随机摆动现象(原因参见第五章反插值法)。

**例 2.14** 用双点弦截法解 $f(x)=xe^x-1=0$。

**解** 取 $x_0=0.5, x_1=0.6$，按式(2.121)计算得

$$x_0=0.5, \qquad f(x_0)=-0.175\ 64$$

$$x_1=0.6, \qquad f(x_1)=0.093\ 27$$

$$x_2=\frac{x_0 f(x_1)-x_1 f(x_0)}{f(x_1)-f(x_0)}=0.565\ 32, \qquad f(x_2)=-0.005\ 03$$

$$x_3=\frac{x_1 f(x_2)-x_2 f(x_1)}{f(x_2)-f(x_1)}=0.567\ 09, \qquad f(x_3)=-0.000\ 15$$

$$x_4=\frac{x_2 f(x_3)-x_3 f(x_2)}{f(x_3)-f(x_2)}=0.567\ 14, \qquad f(x_4)=-0.000\ 01$$

$$x_5=\frac{x_3 f(x_4)-x_4 f(x_3)}{f(x_4)-f(x_3)}=0.567\ 14$$

### 3.5 $|\varphi'(x)|>1$ 的处理方法

在某些情况下，也可以把$|\varphi'(x)|>1$ 的情况改变为收敛的情况。设 $f(x)=0$ 分解为 $x=\varphi(x)$后有

$$|\varphi'(x)|\geqslant k>1, \qquad x\in[a,b] \tag{2.122}$$

这时迭代过程发散。如能将方程 $x=\varphi(x)$右端 $\varphi(x)$中的 $x$ 反解出来得其反函数

$$x=\psi(x) \tag{2.123}$$

则因正、反函数的导数间存在倒数关系得

$$\psi'(x)=\frac{1}{\varphi'(x)} \tag{2.124}$$

$$|\psi'(x)|=\left|\frac{1}{\varphi'(x)}\right|\leqslant\frac{1}{k}<1 \tag{2.125}$$

则按式(2.123)建立迭代公式

$$x_{n+1}=\psi(x_n) \qquad (n=0,1,2,\cdots) \tag{2.126}$$

进行迭代必收敛。

除以上方法外，从式(2.29)可见，要提高收敛速度，还可以使用提高收敛阶数的途径来达到，下面叙述构建高阶迭代公式的具体方法。

## 3.6　高阶迭代函数的构造方法

### 3.6.1　使用反函数台劳公式的构造法

(1) 反函数的导数表达式

设 $y=f(x)$ 的反函数为 $x=\psi(y)$，如直接按反函数表达式求取导数常常会导致复杂的计算。通常我们习惯于正函数的求导工作，所以若能用正函数的导数来表达反函数的导数就可使计算得到简化。反函数的各阶导数与正函数的各阶导数间有如下关系式

$$\psi'(y)=\frac{\mathrm{d}x}{\mathrm{d}y}=\frac{1}{\frac{\mathrm{d}y}{\mathrm{d}x}}=\frac{1}{f'(x)} \tag{2.127}$$

$$\psi''(y)=\frac{\mathrm{d}}{\mathrm{d}y}\frac{1}{f'(x)}=\frac{-f''(x)\frac{\mathrm{d}x}{\mathrm{d}y}}{[f'(x)]^2}=-\frac{f''(x)}{[f'(x)]^2\frac{\mathrm{d}y}{\mathrm{d}x}}=-\frac{f''(x)}{[f'(x)]^3} \tag{2.128}$$

$$\cdots$$

它的一般形式可表为

$$\psi^{(k)}(y)=\frac{X_k}{(y')^{2k-1}}\qquad (k=1,2,\cdots) \tag{2.129}$$

$X_1\sim X_6$ 的表示式在表 2.4 中列出。

**表 2.4**

| | |
|---|---|
| $X_1$ | $1$ |
| $X_2$ | $-y''$ |
| $X_3$ | $-y^{(3)}y'+3[y^{(2)}]^2$ |
| $X_4$ | $-y^{(4)}(y')^2+10y^{(3)}y^{(2)}y'-15[y^{(2)}]^3$ |
| $X_5$ | $-y^{(5)}(y')^3+15y^{(4)}y^{(2)}(y')^2+10[y^{(3)}]^2(y')^2-105y^{(3)}[y^{(2)}]^2y'+105[y^{(2)}]^4$ |
| $X_6$ | $-y^{(6)}(y')^4+21y^{(5)}y^{(2)}(y')^3+35y^{(4)}y^{(3)}(y')^3-210y^{(4)}[y^{(2)}]^2(y')^2-280[y^{(3)}]^2y^{(2)}(y')^2+1260y^{(3)}[y^{(2)}]^3y'-945[y^{(2)}]^5$ |

(2) 二阶迭代函数构造法

对于方程 $f(x)=0$，记 $y=f(x)$，$x=\varphi(y)$，则当 $x=\alpha$ 时，$y=0$。于是有

$$\alpha=\varphi(0)=\varphi(y_n+(0-y_n)) \tag{2.130}$$

将上式在 $y_n$ 点展开为

$$\begin{aligned}\alpha&=\varphi(y_n)+(0-y_n)\varphi'(y_n)+\frac{y_n^2}{2}\varphi''(\eta)\\&=x_n-\frac{f(x_n)}{f'(x_n)}+\frac{[f(x_n)]^2}{2}\cdot\frac{-f''(\xi)}{[f'(\xi)]^3},\qquad \eta\in(0,y_n),\xi\in(\alpha,x_n)\end{aligned} \tag{2.131}$$

在上式的右边截取前两项建立如下迭代公式

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)} \tag{2.132}$$

显见它就是前述的牛顿迭代法。由式(2.131)得

$$\alpha-x_{n+1}=-\frac{[f(x_n)-f(\alpha)]^2}{2}\cdot\frac{f''(\xi)}{[f'(\xi)]^3}$$

$$=-\frac{[f'(\bar{\xi})]^2(x_n-\alpha)^2}{2}\cdot\frac{f''(\xi)}{[f'(\xi)]^3},\qquad \bar{\xi}\in(\alpha,x_n)$$

$$\frac{|x_{n+1}-\alpha|}{|x_n-\alpha|^2}=\left|\frac{[f'(\bar{\xi})]^2f''(\xi)}{2[f'(\xi)]^3}\right| \tag{2.133}$$

如果迭代过程是收敛的，则当 $n\to\infty$ 时， $x_n,\xi,\bar{\xi}\to\alpha$，所以有

$$\frac{|x_{n+1}-\alpha|}{|x_n-\alpha|^2}\xrightarrow{n\to\infty}\frac{|f''(\alpha)|}{2|f'(\alpha)|}=\text{常数} \tag{2.134}$$

从而证得牛顿迭代法是二阶收敛的迭代过程。由式(2.134)可见，当 $f'(\alpha)\approx 0$ 时，牛顿迭代法收敛性较差。

牛顿迭代法收敛的阶数亦可按式(2.30)进行测定。由牛顿迭代法的迭代函数出发，可得以下关系式

$$|\psi'(x)|_{x=\alpha}=\left|\left(x-\frac{f(x)}{f'(x)}\right)'\right|_{x=\alpha}=\left|\frac{f(x)f''(x)}{[f'(x)]^2}\right|_{x=\alpha}=0$$

$$|\psi''(x)|_{x=\alpha}=\left|\frac{-2f(x)[f''(x)]^2+[f'(x)]^2f''(x)+f(x)f'(x)f''(x)}{[f'(x)]^3}\right|_{x=\alpha}$$

$$=\left|\frac{f''(\alpha)}{f'(\alpha)}\right| \tag{2.135}$$

由式(2.135)可见，只要 $f'(\alpha)\neq 0, f''(\alpha)\neq 0$，就有 $|\psi''(\alpha)|\neq 0$，从而证得牛顿迭代法是二阶收敛的。但当 $\alpha$ 为 $f(x)$ 的 $p\geqslant 2$ 重根时，可将 $f(x)$ 表为

$$f(x)=(x-\alpha)^p h(x)\qquad (h(\alpha)\neq 0) \tag{2.136}$$

则有

$$\psi(x)=x-\frac{(x-\alpha)^p h(x)}{h(x)p(x-\alpha)^{p-1}+(x-\alpha)^p h'(x)}$$

$$=x-\frac{(x-\alpha)h(x)}{ph(x)+(x-\alpha)h'(x)}$$

$$\psi'(x)=\frac{h(x)[p(p-1)h(x)+2p(x-\alpha)h'(x)+(x-\alpha)^2h''(x)]}{[ph(x)+(x-\alpha)h'(x)]^2}$$

$$=\frac{(1-\frac{1}{p})+(x-\alpha)\frac{2h'(x)}{ph(x)}+(x-\alpha)^2\frac{h''(x)}{p^2h(x)}}{\left[1+(x-\alpha)\frac{h'(x)}{ph(x)}\right]^2} \tag{2.137}$$

由上式得

$$|\psi'(\alpha)|=|1-\frac{1}{p}| \tag{2.138}$$

因 $|\psi'(\alpha)|<1$，只要 $x_0$ 充分接近 $\alpha$，牛顿迭代法仍收敛，但这时仅有线性收敛性，且 $p$ 愈大，收敛愈慢。

(3) 三阶迭代函数构造法

同法，若将 $\alpha=\varphi(0)=\varphi(y_n+(0-y_n))$ 展开为

$$\alpha=\varphi(y_n)+(0-y_n)\varphi'(y_n)+\frac{y_n^2}{2!}\varphi''(y_n)-\frac{y_n^3}{3!}\varphi'''(\eta),\qquad \eta\in(0,y_n)$$

$$=x_n-\frac{f(x_n)}{f'(x_n)}+\frac{[f(x_n)]^2}{2}\cdot\frac{-f''(x_n)}{[f'(x_n)]^3}-\frac{[f(x_n)]^3}{3!}\varphi'''(\eta)\tag{2.139}$$

截取上式右边前三项建立如下迭代公式

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)}-\frac{[f(x_n)]^2f''(x_n)}{2[f'(x_n)]^3}\tag{2.140}$$

则在迭代过程收敛的情况下有下式成立

$$\frac{|x_{n+1}-\alpha|}{|x_n-\alpha|^3}\xrightarrow{n\to\infty}\frac{1}{3!}\left|[f'(\alpha)]^3\varphi'''(0)\right|=\text{常数}\tag{2.141}$$

证得迭代法(2.140)是三阶收敛的。

仿上推导，可以建立更高阶的迭代函数。

### 3.6.2 使用正函数台劳公式的构造法

高阶迭代函数亦可直接使用 $f(x)$ 台劳展式的部分和来建立，下面叙述具体的构造方法。

(1) 二阶迭代函数构造法

设 $P_1(x)$ 在 $x_n$ 点具有与 $f(x)$ 相同的一、二阶导数，则 $P_1(x)$ 可表为

$$P_1(x)=f(x_n)+\frac{f'(x_n)}{1!}(x-x_n)\tag{2.142}$$

以 $P_1(x)$ 代替 $f(x)$ 求取与横轴的交点设为 $(x_{n+1},0)$ 得

$$0=f(x_n)+\frac{f'(x_n)}{1!}(x_{n+1}-x_n)$$

解得

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)}\tag{2.143}$$

它就是牛顿迭代法。由台劳公式的余式知

$$f(x)-P_1(x)=\frac{f''(\xi_1)}{2!}(x-x_n)^2,\qquad \xi_1\in(x,x_n)$$

$$f(\alpha)-P_1(\alpha)=-P_1(\alpha)=\frac{f''(\xi)}{2!}(\alpha-x_n)^2,\qquad \xi\in(\alpha,x_n)$$

因 $P_1(x_{n+1})=0$，所以有

$$P_1(x_{n+1})-P_1(\alpha)=\frac{f''(\xi)}{2!}(x_n-\alpha)^2$$

$$P'_1(\eta)(x_{n+1}-\alpha)=\frac{f''(\xi)}{2!}(x_n-\alpha)^2,\qquad \eta\in(\alpha,x_{n+1})$$

则下面的极限关系式成立

$$\frac{|x_{n+1}-\alpha|}{|x_n-\alpha|^2}=\left|\frac{f''(\xi)}{2!P'_1(\eta)}\right|\xrightarrow{n\to\infty}\left|\frac{f''(\alpha)}{2f'(\alpha)}\right|=\text{常数}\tag{2.144}$$

从而证得牛顿迭代法是二阶收敛的。

(2) 三阶迭代函数构造法

设 $P_2(x)$ 在 $x_n$ 点具有与 $f(x)$ 相同的一、二、三阶导数，则 $P_2(x)$ 可表为

$$P_2(x)=f(x_n)+\frac{f'(x_n)}{1!}(x-x_n)+\frac{f''(x_n)}{2!}(x-x_n)^2 \tag{2.145}$$

设其与 $x$ 轴的交点为 $(x_{n+1},0)$ 得

$$0=f(x_n)+\frac{f'(x_n)}{1!}(x_{n+1}-x_n)+\frac{f''(x_n)}{2!}(x_{n+1}-x_n)^2$$

这是关于 $(x_{n+1}-x_n)$ 的二次方程，求解得

$$\begin{aligned}x_{n+1}&=x_n+\frac{-f'(x_n)\pm\sqrt{[f'(x_n)]^2-2f(x_n)f''(x_n)}}{f''(x_n)}\\&=x_n-\frac{2f(x_n)}{f'(x_n)\mp\sqrt{[f'(x_n)]^2-2f(x_n)f''(x_n)}}\end{aligned} \tag{2.146}$$

上式中的干号应按 $x_{n+1}$ 与 $x_n$ 最靠近的原则来选定，亦即应使式(2.146)右边第二项的分母绝对值最大原则来选定。这样的迭代公式便是

$$x_{n+1}=x_n-\frac{2f(x_n)}{f'(x_n)+\mathrm{sgn}[f'(x_n)]\sqrt{[f'(x_n)]^2-2f(x_n)f''(x_n)}} \tag{2.147}$$

迭代公式(2.147)的几何意义如图 2.20 所示。

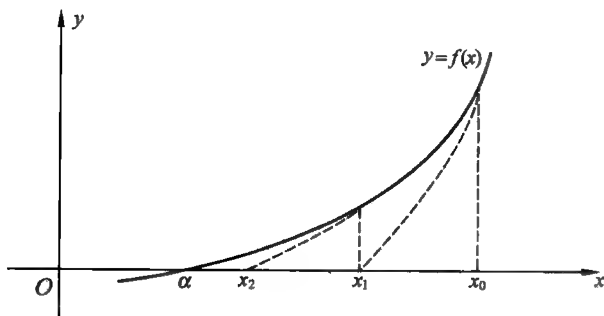



图 2.20

下面推证迭代公式(2.147)收敛的阶数。由台劳公式的余式知

$$f(x)-P_2(x)=\frac{f'''(\xi_1)}{3!}(x-x_n)^3,\qquad \xi_1\in(x,x_n)$$

$$f(\alpha)-P_2(\alpha)=\frac{f'''(\xi)}{3!}(\alpha-x_n)^3,\qquad \xi\in(\alpha,x_n)$$

因 $f(\alpha)=0,P_2(x_{n+1})=0$，所以有

$$P_2(x_{n+1})-P_2(\alpha)=\frac{f'''(\xi)}{3!}(\alpha-x_n)^3$$

$$P'_2(\eta)(x_{n+1}-\alpha)=\frac{f'''(\xi)}{3!}(\alpha-x_n)^3,\qquad \eta\in(\alpha,x_{n+1})$$

因而可得以下关系式

$$\frac{|x_{n+1}-\alpha|}{|x_n-\alpha|^3}=\left|\frac{f'''(\xi)}{6P'_2(\eta)}\right|\xrightarrow{n\to\infty}\left|\frac{f'''(\alpha)}{6f'(\alpha)}\right|=\text{常数} \tag{2.148}$$

从而证得上述迭代法是三阶收敛的。

使用本法建立更高阶迭代函数时，就会导致高次代数方程的求根问题，而高次代数方程求根问题又属于方程求根问题，因而是不可取的。

#### 3.6.3　使用反插值法的构造法

参见第五章 6.2.2 求方程根的反插值法。

## §4　联立方程组的迭代解法

仿单个方程的简单迭代法，对于联立方程组

$$F(\boldsymbol{X})=\begin{bmatrix} f_1(x_1,x_2,\cdots,x_n) \\ f_2(x_1,x_2,\cdots,x_n) \\ \cdots \\ f_n(x_1,x_2,\cdots,x_n) \end{bmatrix}=0 \tag{2.149}$$

可改写为

$$\boldsymbol{X}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}=\begin{bmatrix} \varphi_1(x_1,x_2,\cdots,x_n) \\ \varphi_2(x_1,x_2,\cdots,x_n) \\ \cdots \\ \varphi_n(x_1,x_2,\cdots,x_n) \end{bmatrix}=\Phi(X) \tag{2.150}$$

设 $\boldsymbol{X}^{(0)}=(x_1^{(0)},x_2^{(0)},\cdots,x_n^{(0)})'$ 为根 $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_n)'$ 的零次近似值，则可按以下两种方法进行迭代。

### 4.1　简单迭代法

#### 4.1.1　方法描述

简单迭代法亦称同时迭代法，它由 $\boldsymbol{X}^{(0)}$ 出发，按下式进行迭代

$$\begin{cases} x_1^{(k+1)}=\varphi_1(x_1^{(k)},x_2^{(k)},\cdots,x_n^{(k)}) \\ x_2^{(k+1)}=\varphi_2(x_1^{(k)},x_2^{(k)},\cdots,x_n^{(k)}) \\ \cdots \\ x_n^{(k+1)}=\varphi_n(x_1^{(k)},x_2^{(k)},\cdots,x_n^{(k)}) \end{cases} \tag{2.151}$$

或简记为

$$\boldsymbol{X}^{(k+1)}=\Phi(\boldsymbol{X}^{(k)})$$

式中，$\boldsymbol{X}^{(k)}$ 代表 $\boldsymbol{\alpha}$ 的第 $k$ 次近似值。当 $n\to\infty$ 时，若 $\boldsymbol{X}^{(k)}$ 有极限存在，则该极限即为所求之根 $\boldsymbol{\alpha}$。

当式(2.150)为以下形式

$$x_i=\varphi_i(x_1,x_2,\cdots,x_{i-1},x_{i+1},\cdots,x_n)\quad(i=1,2,\cdots,n) \tag{2.152}$$

时，其对应的简单迭代公式

$$x_i^{(k+1)}=\varphi_i(x_i^{(k)},x_2^{(k)},\cdots,x_{i-1}^{(k)},x_{i+1}^{(k)},\cdots,x_n^{(k)})\quad(i=1,2,\cdots,n) \tag{2.153}$$

称为雅可比迭代公式。

设 $R$ 为含有根 $\boldsymbol{\alpha}$ 的闭域，并且在迭代过程中迭代点留在该闭域中，记

$$\begin{cases} \boldsymbol{X}=(x_1,x_2,\cdots,x_n) \\ a_{ij}=\max\limits_{\boldsymbol{X}\in R}\left|\dfrac{\partial\,\varphi_i(x_1,x_2,\cdots,x_n)}{\partial x_j}\right| \end{cases}$$

则有如下三个收敛的充分条件及其相应的误差估计公式。

充分条件 1　若 $\mu=\max\limits_i \sum\limits_{j=1}^n a_{ij}<1$，则迭代过程收敛，且

$$\max_i |x_i^{(k)}-\alpha_i|\leqslant\frac{\mu^k}{1-\mu}\max_i |x_i^{(1)}-x_i^{(0)}| \tag{2.154}$$

充分条件 2　若 $\nu=\max\limits_j \sum\limits_{i=1}^n a_{ij}<1$，则迭代过程收敛，且

$$\sum_{i=1}^n |x_i^{(k)}-\alpha_i|\leqslant\frac{\nu^k}{1-\nu}\sum_{i=1}^n |x_i^{(1)}-x_i^{(0)}| \tag{2.155}$$

充分条件 3　若 $p=\sqrt{\sum\limits_{i,j=1}^n a_{ij}^2}<1$，则迭代过程收敛，且

$$\sqrt{\sum_{i=1}^n (x_i^{(k)}-\alpha_i)^2}\leqslant\frac{p^k}{1-p}\sqrt{\sum_{i=1}^n (x_i^{(1)}-x_i^{(0)})^2} \tag{2.156}$$

更一般地可建立以下迭代公式

$$\boldsymbol{X}^{(k+1)}=\boldsymbol{X}^{(k)}+wR^{(k)}\quad (w\text{—常数}) \tag{2.157}$$

其中 $R^{(k)}=(R_1^{(k)},R_2^{(k)},\cdots,R_n^{(k)})'$ 为方程(2.150)在 $\boldsymbol{X}=\boldsymbol{X}^{(k)}$ 时的残差。

**例 2.15**　求下列方程组的根

$$\begin{aligned} f_1(x_1,x_2)&=x_1^2-x_2-1=0 \\ f_2(x_1,x_2)&=-x_1+x_2^2-1=0 \end{aligned} \tag{2.158}$$

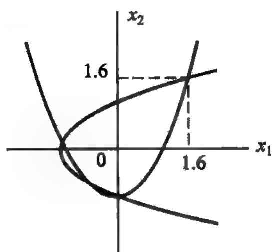



图 2.21

**解**　式(2.158)用方程的形式隐含地给出了 $x_1$ 与 $x_2$ 的对应关系，每一个方程所述及的这种对应关系可用一系列的数值点 $(x_{1i}, x_{2i})(i=1,2,\cdots,n)$ 连成的曲线示于图 2.21 中，其交点所对应的坐标可取为根的近似值。由图 2.21 知其中一个根之初值可取为 $(x_{10},x_{20})=(1.6,1.6)$。

现把方程组(2.158)改写为

$$\begin{cases} x_1=x_2^2-1=\varphi_1(x_1,x_2) \\ x_2=x_1^2-1=\varphi_2(x_1,x_2) \end{cases}$$

则有

$$\frac{\partial\varphi_1}{\partial x_1}=0 \qquad \frac{\partial\varphi_1}{\partial x_2}=2x_2$$

$$\frac{\partial\varphi_2}{2x_1}=2x_1 \qquad \frac{2\varphi_2}{2x_2}=0$$

相应地有

$$a_{11}\approx\left|\frac{\partial\varphi_1}{\partial x_1}\right|_{X^{(0)}}=0,\ a_{12}\approx\left|\frac{\partial\varphi_1}{\partial x_2}\right|_{X^{(0)}}\approx 3.2$$

$$a_{21}\approx\left|\frac{\partial\varphi_2}{\partial x_1}\right|_{X^{(0)}}=3.2,\ a_{22}\approx\left|\frac{\partial\varphi_2}{\partial x_2}\right|_{X^{(0)}}\approx 0$$

如按充分条件 2 判断之

$$\nu_1=a_{11}+a_{21}\approx 3.2$$
$$\nu_2=a_{12}+a_{22}\approx 3.2$$
$$\nu=\max(\nu_1,\nu_2)>1$$

因 $\nu>1$,可能不收敛。如果把方程组重新改写为

$$\begin{cases}x_1=\sqrt{x_2+1}=\varphi_1(x_1,x_2)\\x_2=\sqrt{x_1+1}=\varphi_2(x_1,x_2)\end{cases}\tag{2.159}$$

则

$$\frac{\partial\varphi_1}{\partial x_1}=0,\ \frac{\partial\varphi_1}{\partial x_2}=\frac{1}{2\sqrt{x_2+1}}$$
$$\frac{\partial\varphi_2}{\partial x_1}=\frac{1}{2\sqrt{x_1+1}},\ \frac{\partial\varphi_2}{\partial x_2}=0$$

它们在 $\boldsymbol{X}^{(0)}$ 处的值是

$$\left|\frac{\partial\varphi_1}{\partial x_1}\right|_{\boldsymbol{X}^{(0)}}=0,\ \left|\frac{\partial\varphi_1}{\partial x_2}\right|_{\boldsymbol{X}^{(0)}}=0.31$$
$$\left|\frac{\partial\varphi_2}{\partial x_1}\right|_{\boldsymbol{X}^{(0)}}=0.31,\ \left|\frac{\partial\varphi_2}{\partial x_2}\right|_{\boldsymbol{X}^{(0)}}=0$$

则

$$\nu_1=a_{11}+a_{21}\approx 0.31$$
$$\nu_2=a_{12}+a_{22}\approx 0.31$$
$$\nu=\max(\nu_1,\nu_2)<1$$

因 $\nu<1$,按以下迭代公式进行迭代一定收敛

$$\begin{cases}x_1^{(k+1)}=\sqrt{x_2^{(k)}+1}\\x_2^{(k+1)}=\sqrt{x_1^{(k)}+1}\end{cases}\tag{2.160}$$

计算结果为(1.618,1.618)。

### 4.1.2　关于三个收敛充分条件的证明

**定理 2.8**　若 $\mu=\max\limits_i\sum\limits_{j=1}^{n}a_{ij}<1$,则迭代过程收敛,且

$$\max_i|x_i^{(k)}-\alpha_i|\leqslant\frac{\mu^k}{1-\mu}\max_i|x_i^{(1)}-x_i^{(0)}|$$

**证**　因

$$x_i^{(k)}=\varphi_i(x_1^{(k-1)},x_2^{(k-1)},\cdots,x_n^{(k-1)})$$
$$\alpha_i=\varphi_i(\alpha_1,\alpha_2,\cdots,\alpha_n)$$

两式相减得

$$\begin{aligned}x_i^{(k)}-\alpha_i&=\varphi_i(x_1^{(k-1)},x_2^{(k-1)},\cdots,x_n^{(k-1)})-\varphi_i(\alpha_1,\alpha_2,\cdots,\alpha_n)\\&=\sum_{j=1}^{n}\frac{\partial\bar{\varphi}_i}{\partial x_j}(x_j^{(k-1)}-\alpha_j)\end{aligned}$$

其中　$\dfrac{\partial\bar{\varphi}_i}{\partial x_j}=\dfrac{\partial\varphi_i(\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_n)}{\partial x_j}$,　$\bar{x}_i=\alpha_i+\theta_i(x_i^{(k-1)}-\alpha_i)$　$(0<\theta_i<1)$

则有以下不等式成立

$$|x_i^{(k)}-\alpha_i|\leqslant\sum_{j=1}^{n}\left|\frac{\partial\bar{\varphi}_i}{\partial x_j}\right||x_j^{(k-1)}-\alpha_j|$$
$$\leqslant\left[\sum_{j=1}^{n}\left|\frac{\partial\bar{\varphi}_i}{\partial x_j}\right|\right]\cdot\max_j|x_j^{(k-1)}-\alpha_j|$$
$$\leqslant\mu\max_j|x_j^{(k-1)}-\alpha_j|$$

上式对 $i=1,2,\cdots,n$ 均成立,故下式

$$\max_i|x_i^{(k)}-\alpha_i|\leqslant\mu\max_i|x_j^{(k-1)}-\alpha_j|$$
$$=\mu\max_i|x_i^{(k-1)}-\alpha_i|$$

亦必成立。再令 $k=1,2,\cdots,m$,便得到下列一组不等式

$$\max_i|x_i^{(1)}-\alpha_i|\leqslant\mu\max_i|x_i^{(0)}-\alpha_i|$$
$$\max_i|x_i^{(2)}-\alpha_i|\leqslant\mu\max_i|x_i^{(1)}-\alpha_i|$$
$$\cdots$$
$$\max_i|x_i^{(m)}-\alpha_i|\leqslant\mu\max_i|x_i^{(m-1)}-\alpha_i|$$

将以上各式两边分别相乘并约简得

$$\max_i|x_i^{(m)}-\alpha_i|\leqslant\mu^m\max_i|x_i^{(0)}-\alpha_i|$$

因为 $\mu<1$,所以

$$\lim_{m\to\infty}\max_i|x_i^{(m)}-\alpha_i|\leqslant\lim_{m\to\infty}\mu^m|x_i^{(0)}-\alpha_i|=0$$

即
$$\lim_{m\to\infty}x_i^{(m)}=\alpha_i$$

因此迭代过程是收敛的。由于

$$|x_i^{(m+1)}-x_i^{(m)}|=|\varphi_i(x_1^{(m)},x_2^{(m)},\cdots,x_n^{(m)})-\varphi_i(x_1^{(m-1)},x_2^{(m-1)},\cdots,x_n^{(m-1)})|$$
$$=\left|\sum_{j=1}^{n}\frac{\partial\bar{\varphi}_i}{\partial x_j}(x_j^{(m)}-x_j^{(m-1)})\right|$$
$$\leqslant\mu\max_j|x_j^{(m)}-x_j^{(m-1)}|$$

反复使用上述不等式$(m-1)$次后得

$$|x_i^{(m+1)}-x_i^{(m)}|\leqslant\mu^m\max_j|x_j^{(1)}-x_j^{(0)}|=\mu^m\max_i|x_i^{(1)}-x_i^{(0)}| \tag{2.161}$$

上式对 $i=1,2,\cdots,n$ 均成立,故下式亦必成立

$$\max_i|x_i^{(m+1)}-x_i^{(m)}|\leqslant\mu^m\max_i|x_i^{(1)}-x_i^{(0)}| \tag{2.162}$$

因 $|x_i^{(k+p)}-x_i^{(k)}|\leqslant|x_i^{(k+p)}-x_i^{(k+p-1)}|+|x_i^{(k+p-1)}-x_i^{(k+p-2)}|+\cdots+|x_i^{(k+1)}-x_i^{(k)}|$

推知 $\max_i|x_i^{(k+p)}-x_i^{(k)}|\leqslant\max_i|x_i^{(k+p)}-x_i^{(k+p-1)}|+\max_i|x_i^{(k+p-1)}-x_i^{(k+p-2)}|+\cdots+\max_i|x_i^{(k+1)}-x_i^{(k)}|$

$$\leqslant(\mu^{k+p-1}+\mu^{k+p-2}+\cdots+\mu^k)\max_i|x_i^{(1)}-x_i^{(0)}|$$
$$\leqslant\frac{\mu^k}{1-\mu}\max_i|x_i^{(1)}-x_i^{(0)}| \tag{2.163}$$

上式当 $p\to\infty$时,即证得结果式(2.154)。

**定理 2.9** 若 $\nu=\max\limits_j\sum\limits_{i=1}^{n}a_{ij}<1$,则迭代过程收敛,且

$$\sum_{i=1}^{n}|x_i^{(k)}-\alpha_i|\leqslant\frac{\nu^k}{1-\nu}\sum_{i=1}^{n}|x_i^{(1)}-x_i^{(0)}|$$

**证**　因

$$|x_i^{(k)}-\alpha_i|\leqslant\sum_{j=1}^{n}\left|\frac{\partial\bar{\varphi}_i}{\partial x_j}\right||x_j^{(k-1)}-\alpha_j|$$

两边求和得

$$\begin{aligned}\sum_{i=1}^{n}|x_i^{(k)}-\alpha_i|&\leqslant\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\frac{\partial\bar{\varphi}_i}{\partial x_j}\right||x_j^{(k-1)}-\alpha_j|\\&=\sum_{j=1}^{n}\sum_{i=1}^{n}\left|\frac{\partial\bar{\varphi}_i}{\partial x_j}\right||x_j^{(k-1)}-\alpha_j|\\&\leqslant\sum_{j=1}^{n}\nu|x_j^{(k-1)}-\alpha_j|\\&=\nu\sum_{j=1}^{n}|x_j^{(k-1)}-\alpha_j|\\&=\nu\sum_{i=1}^{n}|x_i^{(k-1)}-\alpha_i|\end{aligned}\tag{2.164}$$

令 $k=1,2,\cdots,m$，得下列不等式

$$\sum_{i=1}^{n}|x_i^{(1)}-\alpha_i|\leqslant\nu\sum_{i=1}^{n}|x_i^{(0)}-\alpha_i|$$

$$\sum_{i=1}^{n}|x_i^{(2)}-\alpha_i|\leqslant\nu\sum_{i=1}^{n}|x_i^{(1)}-\alpha_i|$$

$$\cdots$$

$$\sum_{i=1}^{n}|x_i^{(m)}-\alpha_i|\leqslant\nu\sum_{i=1}^{n}|x_i^{(m-1)}-\alpha_i|$$

将以上各式两边分别相乘并约简得

$$\sum_{i=1}^{n}|x_i^{(m)}-\alpha_i|\leqslant\nu^m\sum_{i=1}^{n}|x_i^{(0)}-\alpha_i|\tag{2.165}$$

因 $\nu<1$，所以有

$$\lim_{m\to\infty}\sum_{i=1}^{n}|x_i^{(m)}-\alpha_i|\leqslant\lim_{m\to\infty}\nu^m\sum_{i=1}^{n}|x_i^{(0)}-\alpha_i|=0$$

即

$$\lim_{m\to\infty}x_i^{(m)}=\alpha_i$$

亦即迭代过程是收敛的。

仿前可证得误差估计公式(2.155)。

**定理 2.10**　若 $p=\sqrt{\sum_{i,j=1}^{n}a_{ij}^2}<1$，则迭代过程收敛，且

$$\sqrt{\sum_{i=1}^{n}(x_i^{(k)}-\alpha_i)^2}\leqslant\frac{p^k}{1-p}\sqrt{\sum_{i=1}^{n}(x_i^{(1)}-x_i^{(0)})^2}$$

**证**　因

$$x_i^{(k)}-\alpha_i=\sum_{j=1}^{n}\frac{\partial\bar{\varphi}_i}{\partial x_j}(x_j^{(k-1)}-\alpha_j)$$

将上式两边平方并求和得

$$\sum_{i=1}^{n}(x_i^{(k)}-\alpha_i)^2=\sum_{i=1}^{n}\left[\sum_{j=1}^{n}\frac{\partial\bar{\varphi}_i}{\partial x_j}(x_j^{(k-1)}-\alpha_j)\right]^2 \tag{2.166}$$

利用不等式

$$\left(\sum_{s=1}^{n}a_sb_s\right)^2\leqslant\left(\sum_{s=1}^{n}a_s^2\right)\cdot\left(\sum_{s=1}^{n}b_s^2\right)$$

可将式(2.166)化为

$$\begin{aligned}\sum_{i=1}^{n}(x_i^{(k)}-\alpha_i)^2&\leqslant\sum_{i=1}^{n}\left[\sum_{j=1}^{n}\left(\frac{\partial\bar{\varphi}_i}{\partial x_j}\right)^2\right]\cdot\left[\sum_{j=1}^{n}(x_j^{(k-1)}-\alpha_j)^2\right]\\&=\left[\sum_{i,j=1}^{n}\left(\frac{\partial\bar{\varphi}_i}{\partial x_j}\right)^2\right]\cdot\left[\sum_{j=1}^{n}(x_j^{(k-1)}-\alpha_j)^2\right]\\&\leqslant p^2\sum_{j=1}^{n}(x_j^{(k-1)}-\alpha_j)^2\\&=p^2\sum_{i=1}^{n}(x_i^{(k-1)}-\alpha_i)^2\end{aligned} \tag{2.167}$$

令 $k=1,2,\cdots,m$ 得

$$\sum_{i=1}^{n}(x_i^{(1)}-\alpha_i)^2\leqslant p^2\sum_{i=1}^{n}(x_i^{(0)}-\alpha_i)^2$$

$$\sum_{i=1}^{n}(x_i^{(2)}-\alpha_i)^2\leqslant p^2\sum_{i=1}^{n}(x_i^{(1)}-\alpha_i)^2$$

$$\cdots$$

$$\sum_{i=1}^{n}(x_i^{(m)}-\alpha_i)^2\leqslant p^2\sum_{i=1}^{n}(x_i^{(m-1)}-\alpha_i)^2$$

将上述各式两边分别相乘并约简后得

$$\sum_{i=1}^{n}(x_i^{(m)}-\alpha_i)^2\leqslant p^{2m}\sum_{i=1}^{n}(x_i^{(0)}-\alpha_i)^2$$

因 $p<1$，对上式两边取极限得

$$\lim_{m\to\infty}\sum_{i=1}^{n}(x_i^{(m)}-\alpha_i)^2\leqslant\lim_{m\to\infty}p^{2m}\sum_{i=1}^{n}(x_i^{(0)}-\alpha_i)^2=0$$

所以必有

$$\lim_{m\to\infty}x_i^{(m)}=\alpha_i$$

仿前，同样可证得误差估计公式(2.156)。

## 4.2 赛德尔迭代法

赛德尔迭代法与简单迭代法在迭代方式上的不同处在于前者在获得了新的近似值以后，用它们替代旧近似值来参与以后的相继迭代计算，因而得到以下迭代公式

$$\begin{cases}x_1^{(k+1)}=\varphi_1(x_1^{(k)},x_2^{(k)},\cdots,x_n^{(k)})\\x_2^{(k+1)}=\varphi_2(x_2^{(k+1)},x_2^{(k)},\cdots,x_n^{(k)})\\x_3^{(k+1)}=\varphi_3(x_1^{(k+1)},x_2^{(k+1)},x_3^{(k)},\cdots,x_n^{(k)})\\\cdots\\x_n^{(k+1)}=\varphi_n(x_1^{(k+1)},x_2^{(k+1)},\cdots,x_{n-1}^{(k+1)},x_n^{(k)})\end{cases} \tag{2.168}$$

特别当分解式为式(2.152)时所对应的赛德尔迭代法

$$x_i^{(k+1)}=\varphi_i(x_1^{(k+1)},\cdots,x_{i-1}^{(k+1)},x_{i+1}^{(k)},\cdots,x_n^{(k)})\ (i=1,2,\cdots,n) \tag{2.169}$$

称为高斯-赛德尔迭代法。

**例 2.16** 用赛德尔迭代法求解例 2.15。

**解**:我们按分解式(2.159)建立赛德尔迭代公式如下

$$\begin{cases}x_1^{(k+1)}=\sqrt{x_2^{(k)}+1}\\ x_2^{(k+1)}=\sqrt{x_1^{(k+1)}+1}\end{cases} \tag{2.170}$$

取 $\boldsymbol{X}^{(0)}=(1.6,1.6)'$,按式(2.170)计算得

$$\begin{cases}x_1^{(1)}=\sqrt{1.6+1}=1.612\\ x_2^{(1)}=\sqrt{1.612+1}=1.616\end{cases},\ \begin{cases}x_1^{(2)}=\sqrt{1.616+1}=1.617\\ x_2^{(2)}=\sqrt{1.617+1}=1.618\end{cases},\ \begin{cases}x_1^{(3)}=\sqrt{1.618+1}=1.618\\ x_2^{(3)}=\sqrt{1.618+1}=1.618\end{cases}$$

从以上赛德尔迭代法可见,赛德尔迭代法把每次计算出来的新值代替旧值,因此它只需一组存储单元用于存放近似值。而在简单迭代法中,需要两组存储单元存放新、旧近似值。另外,在算法上的这种改变给编程带来方便,使程序设计更加简易。

## 4.3 松弛迭代法

首先说明松弛一词的含义,对于 $F(\boldsymbol{X})=0$,设 $R_i^{(k)}$ 为其第 $i$ 个方程当 $\boldsymbol{X}=\boldsymbol{X}^{(k)}$ 时的残差。一般说来,它不等于零,否则该近似值 $\boldsymbol{X}^{(k)}$ 可能是解了。现修改 $\boldsymbol{X}^{(k)}$ 的部分或全部变量值为新的近似值 $\boldsymbol{X}^{(k+1)}$,使该方程的残差 $R_i^{(k+1)}=0$,这时我们就说该方程被松弛了或被削弱了。如果方程 $F(\boldsymbol{X})=0$ 的全部残差都同时被松弛为零时,则所对应的近似值就是所求的根 $\alpha$ 了。

根据上述松弛法的原理,我们需要确定 $\boldsymbol{X}^{(k)}$ 中被修改变量的数量以及 $F(\boldsymbol{X})=0$ 中方程被松弛的顺序。为降低求解方程的复杂性,最简单的做法就是只改变 $\boldsymbol{X}^{(k)}$ 中的一个变量 $x_i^{(k)}$ 的数值为 $\hat{x}_i$,使某个方程的残差为零,这时松弛方程的问题就转化为对一个只含有一个未知量 $\hat{x}_i$ 的方程求解问题。为了对变量有规律地进行修改,一般可约定第 $i$ 个方程总是改变其第 $i$ 个变量的数值,使该方程的残差为零。至于方程的松弛顺序可以采用不同的控制策略,相应地就产生了不同的松弛迭代法,以下我们介绍一种按方程排列顺序相继地进行松弛的方法即逐次松弛迭代法。

### 4.3.1 简单迭代方式下的逐次松弛法

设 $\boldsymbol{X}^{(k)}=(x_1^{(k)},x_2^{(k)},\cdots,x_n^{(k)})'$是 $\alpha$ 的 $k$ 次近似值,由 $i=1$ 开始,改变 $x_i^{(k)}$ 为 $\hat{x}_i$ 使方程

$$f_i(x_1^{(k)},x_2^{(k)},\cdots,x_{i-1}^{(k)},\hat{x}_i,x_{i+1}^{(k)},\cdots,x_n^{(k)})\equiv 0\ (i=1,2,\cdots,n) \tag{2.171}$$

由上述方程得到的解 $\hat{x}_i$ 记作 $x_i^{(k+1)}$,继续进行下去,直至 $i=n$ 为止,由此就可获得新的近似值 $\boldsymbol{X}^{(k+1)}=(x_1^{(k+1)},x_2^{(k+1)},\cdots,x_n^{(k+1)})'$。以上过程重复至全部修改量满足精度要求为止,则其最终的近似值就是方程 $F(\boldsymbol{X})=0$ 的解。因式(2.171)是一个单变量非线性方程,其解 $\hat{x}_i$ 亦可采用迭代法求解之。这样就形成了双层迭代,主迭代为松弛迭代法;求解式(2.171)中方程的迭代法则为内层迭代法。当由式(2.171)可直接地解得 $\hat{x}_i$ 为

$$\hat{x}_i=\varphi_i(x_1^{(k)},\cdots,x_{i-1}^{(k)},x_{i+1}^{(k)},\cdots,x_n^{(k)})\ (i=1,2,\cdots,n) \tag{2.172}$$

令 $x_i^{(k+1)}=\hat{x}_i$,则上述的逐次松弛迭代法完全等同于雅可比迭代法(2.153)。

### 4.3.2 赛德尔迭代方式下的逐次松弛法

在本法中,被松弛的方程的形式为

$$f_i(x_1^{(k+1)},x_2^{(k+1)},\cdots,x_{i-1}^{(k+1)},\hat{x}_i,x_{i+1}^{(k)},\cdots,x_n^{(k)})\equiv 0\ (i=1,2,\cdots,n) \tag{2.173}$$

由上述方程得到的解 $\hat{x}_i$ 记作 $x_i^{(k+1)}(i=1,2,\cdots,n)$，重复地进行下去直至全部修改量满足精度要求为止。当式(2.173)中的 $\hat{x}_i$ 可以直接解出为

$$\hat{x}_i=\varphi_i(x_i^{(k+1)},\cdots,x_{i-1}^{(k+1)},x_i^{(k)},\cdots,x_n^{(k)})\ (i=1,2,\cdots,n) \tag{2.174}$$

令 $x^{(k+1)}=\hat{x}_i$，则上述迭代法完全等同于高斯-赛德尔迭代法(2.169)。

#### 4.3.3 对具有最大残差绝对值的方程实施松弛的迭代法

方程的松弛顺序亦可按残差绝对值为最大的原则来确定，具体过程如下：

对于 $\boldsymbol{X}^{(k)}=(x_1^{(k)},x_2^{(k)},\cdots,x_n^{(k)})'$，计算出 $R^{(k)}=(R_1^{(k)},R_2^{(k)},\cdots,R_n^{(k)})'$，设其中残差绝对值最大者为 $|R_i^{(k)}|=\max\limits_l|R_l^{(k)}|(l=1,2,\cdots,n)$，由此确定出对第 $i$ 个方程进行松弛，即改变 $x_i^{(k)}$ 为 $\hat{x}_i$ 使第 $i$ 个方程松弛为零

$$f_i(x_1^{(k)},\cdots,x_{i-1}^{(k)},\hat{x}_i,x_{i+1}^{(k)},\cdots,x_n^{(k)})\equiv 0 \tag{2.175}$$

由式(2.175)解得 $\hat{x}_i$ 后，令新的近似值为 $\boldsymbol{X}^{(k+1)}=(x_1^{(k)},\cdots,x_{i-1}^{(k)},\hat{x}_i,x_{i+1}^{(k)},\cdots,x_n^{(k)})'$，使用 $\boldsymbol{X}^{(k+1)}$，计算新的残差 $R^{(k+1)}=(R_1^{(k+1)},R_2^{(k+1)},\cdots,R_n^{(k+1)})'$，重复以上过程直至全部修改量满足精度要求为止，即可取得最终近似值为方程 $F(\boldsymbol{X})=0$ 的解。

#### 4.3.4 带有松弛因子的松弛迭代法

在本法中，对由式(2.171)、式(2.173)、式(2.175)解得的 $\hat{x}_i$，使用以下公式来确定新的近似值

$$x_i^{(k+1)}=x_i^{(k)}+\widetilde{w}(\hat{x}_i-x_i^{(k)})\ (i=1,2,\cdots,n) \tag{2.176}$$

其中 $\widetilde{w}$ 为引入的参数，称为松弛因子，采用(2.176)式对变量进行修改的松弛法称为带有松弛因子的松弛法。当 $\widetilde{w}=1$ 时，所得 $\hat{x}_i$ 可使第 $i$ 个方程完全松弛为零，称为恰好松弛法。对于 $\widetilde{w}>1$ 或 $\widetilde{w}<1$，则第 $i$ 个方程可能被松弛过头或松弛得不够，分别称它们为超松弛法($\widetilde{w}>1$)和低松弛法($\widetilde{w}<1$)。

## §5 联立方程组的牛顿解法

与单个方程的牛顿解法

$$\begin{cases}\Delta x_k=x_{k+1}-x_k=-\dfrac{f(x_k)}{f'(x_k)}\\ f'(x_k)\cdot\Delta x_k=-f(x_k)\end{cases} \tag{2.177}$$

相类比，在联立方程组 $F(X)=0$ 的情况下，可建立如下牛顿迭代公式

$$\begin{cases}F'(\boldsymbol{X}^{(k)})\cdot\Delta\boldsymbol{X}^{(k)}=-F(\boldsymbol{X}^{(k)})\\ \boldsymbol{X}^{(k+1)}=\boldsymbol{X}^{(k)}+\Delta\boldsymbol{X}^{(k)}\end{cases} \tag{2.178}$$

其中

$$F(\boldsymbol{X}^{(k)})=\begin{bmatrix}f_1(x_1^{(k)},x_2^{(k)},\cdots,x_n^{(k)})\\ f_2(x_1^{(k)},x_2^{(k)},\cdots,x_n^{(k)})\\ \cdots\\ f_n(x_1^{(k)},x_2^{(k)},\cdots,x_n^{(k)})\end{bmatrix}=\begin{bmatrix}f_1\\ f_2\\ \vdots\\ f_n\end{bmatrix}_{\boldsymbol{X}^{(k)}}$$

$$F'(\boldsymbol{X}^{(k)})=\begin{bmatrix}\dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n}\\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_n}\\ & \cdots & & \\ \dfrac{\partial f_n}{\partial x_1} & \dfrac{\partial f_n}{\partial x_2} & \cdots & \dfrac{\partial f_n}{\partial x_n}\end{bmatrix}_{\boldsymbol{X}^{(k)}} \qquad \Delta\boldsymbol{X}^{(k)}=\begin{bmatrix}\Delta x_1^{(k)}\\ \Delta x_2^{(k)}\\ \vdots\\ \Delta x_n^{(k)}\end{bmatrix}$$

式(2.178)是关于 $\Delta x_j^{(k)}(j=1,2,\cdots,n)$ 的线性方程组

$$\begin{cases}\left(\dfrac{\partial f_i}{\partial x_1}\right)_{\boldsymbol{X}^{(k)}}\Delta x_1^{(k)}+\left(\dfrac{\partial f_i}{\partial x_2}\right)_{\boldsymbol{X}^{(k)}}\Delta x_2^{(k)}+\cdots+\left(\dfrac{\partial f_i}{\partial x_n}\right)_{\boldsymbol{X}^{(k)}}\Delta x_n^{(k)}=-(f_i)_{\boldsymbol{X}^{(k)}}\\ x_i^{(k+1)}=x_i^{(k)}+\Delta x_i^{(k)} \quad (i=1,2,\cdots,n;k=0,1,2,\cdots)\end{cases} \tag{2.179}$$

当系数行列式 $|F'(\boldsymbol{X}^{(k)})|\neq0$ 时，方程组有唯一解 $\Delta x_j^{(k)}(j=1,2,\cdots,n)$。在获得新的近似值 $\boldsymbol{X}^{(k+1)}$ 后，重复上述过程直到

$$|\Delta x_j^{(k)}|<\varepsilon(\text{给定误差要求}) \qquad (j=1,2,\cdots,n)$$

满足为止。按式(2.179)进行迭代，每次都要解一个线性方程组，并且其系数行列式每一次也不同，这个方法就称为牛顿迭代法，它具有二阶收敛速度。牛顿迭代法对初值的精度要求较高，当初值选取得较为精确时，收敛是很快的。

类似地，求解 $F(X)=0$ 的牛顿法同样可以变形地使用。如果将式(2.179)中的 $\left(\dfrac{\partial f_i}{\partial x_j}\right)_{\boldsymbol{X}^{(k)}}$ 用固定值 $\left(\dfrac{\partial f_i}{\partial x_j}\right)_{\boldsymbol{X}^{(0)}}$ 代替，则式(2.179)就转化为联立方程组的简化牛顿法。为提高收敛速度，可采用修正的牛顿法，具体处理如下：取 $\boldsymbol{X}^{(0)}$，使用 $F'(\boldsymbol{X}^{(0)})$ 作 $m$ 次简化牛顿法得 $\boldsymbol{X}^{(1)}$，再以 $\boldsymbol{X}^{(1)}$ 代替 $\boldsymbol{X}^{(0)}$，使用 $F'(\boldsymbol{X}^{(1)})$ 作 $m$ 次简化牛顿法得 $\boldsymbol{X}^{(2)}$，如此推作，直至最终近似值达到精度要求为止。

为扩大收敛范围，可将下山法融合到牛顿法中构成牛顿下山法

$$F'(\boldsymbol{X}^{(k)})\cdot\Delta\boldsymbol{X}^{(k)}=-\lambda_i F(\boldsymbol{X}^{(k)}) \quad (0<\lambda_i<1)$$

$$\|F(\boldsymbol{X}^{(k+1)})\|<\|F(\boldsymbol{X}^{(k)})\| \quad (\text{下山条件}) \quad k=0,1,2,\cdots \tag{2.180}$$

式中 $\|\cdot\|$ 为范数，参数 $\lambda_i$ 据下山条件来选取，其计算过程的处理方法与单个方程的牛顿下山法类同。

**例 2.17** 试用牛顿迭代法求解下列方程组的根。

$$\begin{cases}f_1(x,y)=2x^3-y^2-1=0\\ f_2(x,y)=xy^3-y-4=0\\ (x^{(0)},y^{(0)})=(1.2,1.7)\end{cases}$$

**解** 因为

$$\frac{\partial f_1}{\partial x}=6x^2,\quad \frac{\partial f_1}{\partial y}=-2y$$

$$\frac{\partial f_2}{\partial x}=y^3,\quad \frac{\partial f_2}{\partial y}=3xy^2-1$$

所以得 $$\frac{\partial f_1(x^{(0)},y^{(0)})}{\partial x}=8.64,\quad \frac{\partial f_1(x^{(0)},y^{(0)})}{\partial y}=-3.40,\quad f_1(x_0,y_0)=-0.434$$

$$\frac{\partial f_2(x^{(0)},y^{(0)})}{\partial x}=4.91,\quad \frac{\partial f_2(x^{(0)},y^{(0)})}{\partial y}=9.4,\quad f_2(x_0,y_0)=0.195\,6$$

于是按式(2.179)可建立下面的线性方程组

$$\begin{cases}8.64\Delta x_1^{(0)}-3.40\Delta x_2^{(0)}=0.434\\4.91\Delta x_1^{(0)}+9.4\Delta x_2^{(0)}=-0.1956\end{cases}$$

解得

$$\Delta x_1^{(0)}=0.0349,\quad \Delta x_2^{(0)}=-0.0390$$

则得根的一次近似值为

$$\begin{cases}x^{(1)}=x^{(0)}+\Delta x_1^{(0)}=1.2+0.0349=1.2349\\y^{(1)}=y^{(0)}+\Delta x_2^{(0)}=1.7+(-0.0390)=1.6610\end{cases}$$

再进行一次同样计算过程得

$$\begin{cases}x^{(2)}=1.2343\\y^{(2)}=1.6615\end{cases}$$

# §6 联立方程组的延拓解法

对于大多数迭代法，一般都是局部收敛的方法，即要求初值充分接近于根，才能使迭代序列收敛于根。实际计算中要找到满足要求的初值往往是件很困难的事情。下面介绍一种延拓法，它可以看做是一种扩大收敛域的求根方法。

## 6.1 同伦方程组及其建立方法

对于联立方程组 $\boldsymbol{F}(\boldsymbol{X})=0$，用延拓法求解联立方程组，首先要建立同伦方程组 $\boldsymbol{H}(\boldsymbol{X},t)=0$，这里 $\boldsymbol{X}=(x_1,x_2,\cdots,x_n)$，$t$ 为引入的参数。同伦方程组是满足以下条件的联立方程组。

(1) $\boldsymbol{H}(\boldsymbol{X}^{(0)},0)\equiv 0$ (2.181)

式中，$\boldsymbol{X}^{(0)}=(x_1^{(0)},x_2^{(0)},\cdots,x_n^{(0)})$是任意取定的一组初值。式(2.181)表明，当 $t=0$ 时，同伦方程组为具有已知解 $\boldsymbol{X}^{(0)}$ 的联立方程组。

(2) $\boldsymbol{H}(\boldsymbol{X},1)=\boldsymbol{F}(\boldsymbol{X})=0$ (2.182)

上式表明，当 $t=1$ 时，同伦方程组与原联立方程组 $\boldsymbol{F}(\boldsymbol{X})=0$ 全同。

满足以上条件(1)、(2)的同伦方程组可以采用不同的方法构造出来。例如

$$\begin{cases}\boldsymbol{H}(\boldsymbol{X},t)=t\boldsymbol{F}(\boldsymbol{X})+(1-t)\boldsymbol{\Phi}(\boldsymbol{X})\\\boldsymbol{\Phi}(\boldsymbol{X})=\boldsymbol{F}(\boldsymbol{X})-\boldsymbol{F}(\boldsymbol{X}^{(0)})\end{cases}\tag{2.183}$$

或

$$\boldsymbol{H}(\boldsymbol{X},t)=\boldsymbol{F}(\boldsymbol{X})-(1-t)\boldsymbol{F}(\boldsymbol{X}^{(0)})\tag{2.184}$$

或者将条件(2)改成当 $t\to\infty$时，$\boldsymbol{H}(\boldsymbol{X},t)\longrightarrow\boldsymbol{F}(\boldsymbol{X})$，可以构造

$$\begin{cases}\boldsymbol{H}(\boldsymbol{X},t)=\boldsymbol{F}(\boldsymbol{X})-\mathrm{e}^{-t}\boldsymbol{F}(\boldsymbol{X}^{(0)})=0,\qquad t\in[0,\infty)\\\boldsymbol{H}(\boldsymbol{X},0)=\boldsymbol{F}(\boldsymbol{X})-\boldsymbol{F}(\boldsymbol{X}^{(0)})\\\boldsymbol{H}(\boldsymbol{X},t)\xrightarrow{t\to\infty}\boldsymbol{F}(\boldsymbol{X})\end{cases}\tag{2.185}$$

等。

## 6.2 求解方法

由上可见，求解联立方程解的问题可以转化为求解同伦方程组 $\boldsymbol{H}(\boldsymbol{X},1)=0$ 的解的问题。

为了求取 $\boldsymbol{H}(\boldsymbol{X},1)=0$ 的解，延拓法的思想就是从已知解 $\boldsymbol{X}^{(0)}$ 出发，逐步引渡到 $\boldsymbol{F}(\boldsymbol{X})=0$ 的未知解，具体方法如下。

首先将 $t$ 的值域 $t\in[0,1]$ 等距或不等距地分划为

$$0=t_0<t_1<t_2<\cdots<t_{N-1}<t_N=1 \tag{2.186}$$

建立相应的同伦方程组

$$\boldsymbol{H}_i=\boldsymbol{H}(\boldsymbol{X},t_i)=0 \qquad (i=1,2,\cdots,N) \tag{2.187}$$

以下取 $\boldsymbol{X}^{(0)}$ 作为 $\boldsymbol{H}_1=0$ 的初值，用某种迭代法求解出 $\boldsymbol{H}_1=0$ 的解 $\boldsymbol{X}^{(1)}$。再以 $\boldsymbol{X}^{(1)}$ 作为 $\boldsymbol{H}_2=0$ 的初值，求解出 $\boldsymbol{H}_2=0$ 的解 $\boldsymbol{X}^{(2)}$，如此推导下去，直至求出 $\boldsymbol{H}_N=0$ 的解 $\boldsymbol{X}^{(N)}$ 为止。在上述求解过程中，如果 $t_i-t_{i-1}$ 充分小，可以期望 $\boldsymbol{X}^{(i-1)}$ 是 $\boldsymbol{X}^{(i)}$ 的一个足够好的近似，从而使用局部收敛的迭代法就可获得收敛的计算结果，这就是延拓法的基本思想。当上述同伦方程组的解 $\boldsymbol{X}^{(i)}$ $(i=1,2,\cdots,N)$ 为 $t$ 的连续函数时，则上述延拓过程是可实现的，而 $\boldsymbol{X}^{(N)}$ 就是原方程 $\boldsymbol{F}(\boldsymbol{X})=0$ 的解。

**例 2.18**　用延拓法求下列方程组的根

$$\begin{cases} f_1(x,y)=x+3\lg x-y^2=0 \\ f_2(x,y)=2x^2-xy-5x+1=0 \end{cases} \tag{2.188}$$

**解**　我们将式(2.188)中的 $f_1(x,y)=0$ 与 $f_2(x,y)=0$ 的曲线示于图 2.22 中，它们有两个交点，即联立方程组(2.188)具有两个根，其中一根位于 $(x,y)=(3.4,2.2)$ 附近。如取初值为 $(x_0,y_0)=(10,10)$，按(2.183)式建立同伦方程组

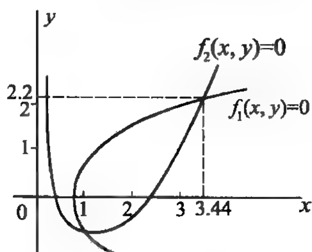



图 2.22

$$\boldsymbol{\Phi}(\boldsymbol{X})=\begin{cases} \Phi_1(X)=f_1(x,y)-f_1(10,10)=x+3\lg x-y^2+87 \\ \Phi_2(X)=f_2(x,y)-f_2(10,10)=2x^2-xy+1-51 \end{cases}$$

$$\boldsymbol{H}(\boldsymbol{X},t)=\begin{cases} H_1(X,t)=tf_1(x,y)+(1-t)\Phi_1(X)=x+3\lg x-y^2+87(1-t)=0 \\ H_2(X,t)=tf_2(x,y)+(1-t)\Phi_2(X)=2x^2-xy-5x+1-51(1-t)=0 \end{cases} \tag{2.189}$$

将式(2.189)中的 $\boldsymbol{H}(\boldsymbol{X},t)=0$ 改写为如下的迭代公式

$$\begin{cases} x_{n+1}=\sqrt{\dfrac{x_n(y_n+5)-1+51(1-t_i)}{2}} \\ y_{n+1}=\sqrt{x_n+3\lg x_n+87(1-t_i)} \end{cases} \tag{2.190}$$

今取 $t_i(i=0,1,2,3,4)$ 为

$$t_0=0,t_1=0.25,t_2=0.5,t_3=0.75,t_4=1$$

对应的迭代公式及按延拓法所求得的解为

$t_1=0.25$：

$$\begin{cases} x_{n+1}=\sqrt{\dfrac{x_n(y_n+5)+37.25}{2}} \\ y_{n+1}=\sqrt{x_n+3\lg x_n+65.25} \end{cases}$$

初值 $\boldsymbol{X}^{(0)}=(x_0,y_0)=(10,10)$

解 $\boldsymbol{X}^{(1)}=(x_1,y_1)=(9,9)$

$t_2=0.50$：

$$\begin{cases} x_{n+1}=\sqrt{\dfrac{x_n(y_n+5)+24.5}{2}} \\ y_{n+1}=\sqrt{x_n+3\lg x_n+43.5} \end{cases}$$

初值 $\boldsymbol{X}^{(1)}=(9,9)$

解 $\boldsymbol{X}^{(2)}=(x_2,y_2)=(8,7)$

$t_3=0.75$:

$$\begin{cases} x_{n+1}=\sqrt{\dfrac{x_n(y_n+5)+11.75}{2}} \\ y_{n+1}=\sqrt{x_n+3\lg x_n+21.75} \end{cases}$$

初值 $\boldsymbol{X}^{(2)}=(8,7)$

解 $\boldsymbol{X}^{(3)}=(x_3,y_3)=(7,6)$

$t_4=1.00$:

$$\begin{cases} x_{n+1}=\sqrt{\dfrac{x_n(y_n+5)-1}{2}} \\ y_{n+1}=\sqrt{x_n+3\lg x_n} \end{cases}$$

初值 $\boldsymbol{X}^{(3)}=(7,6)$

解 $\boldsymbol{X}^{(4)}=(x_4,y_4)=(3.487,2.262)$

# 习 题 二

2.1 用对分法求方程 $e^x+10^x-2=0$ 在(0,1)内的根，其绝对误差限 $\varepsilon=10^{-3}$。

2.2 求方程 $x^3-x^2-1=0$ 在 $x_0=1.5$ 附近的根，设将方程改写成下列等价形式，并建立相应的迭代公式：

(1) $x=1+\dfrac{1}{x^2}$，迭代公式 $x_{n+1}=1+\dfrac{1}{x_n^2}$

(2) $x^3=1+x^2$，迭代公式 $x_{n+1}=\sqrt[3]{1+x_n^2}$

(3) $x^2=\dfrac{1}{x-1}$，迭代公式 $x_{n+1}=\dfrac{1}{\sqrt{x_n-1}}$

试分析每种迭代公式的收敛性。

2.3 用埃特肯法求方程 $x^3-x^2-1=0$ 在 $x_0=1.5$ 附近的根，$\varepsilon=10^{-4}$。

2.4 分别用牛顿迭代法、弦截法求方程 $x^3-x^2-x-1=0$ 的正根，$\varepsilon=10^{-4}$。

2.5 用下列方法求方程 $x^3-3x-1=0$ 在 $x=2$ 附近的根，准确到四位有效数字。

(1)迭代法；(2)牛顿迭代法；(3)弦截法，$x_0=2$，$x_1=1.9$。

2.6 用迭代法求方程组

$$\begin{cases}x^2+y^2-1=0\\x^2-y=0\end{cases}$$

在 $x_0=0.8$，$y_0=0.6$ 附近的根，准确到三位小数。

2.7 用牛顿迭代法解方程组

$$\begin{cases}x+2y-3=0\\2x^2+y^2-5=0\end{cases}$$

初值取 $(x^{(0)},y^{(0)})=(-1,2)$，根的近似值准确到三位小数。

2.8 将牛顿迭代应用于 $x^p-c=0$，导出求正数 $c$ 的 $p$ 次根的迭代公式。

2.9 在 $|x|<1$，$|y|<1$ 域内，用迭代法求解下述方程组

$$\begin{cases}2x-\cos y=0\\2y-\sin x=0\end{cases}$$

要求根的近似值准确到两位小数。

2.10 方程 $f(x)=(x-1)^8=0$，应用切线法时，若取 $x_0=1.1$ 进行计算，发现收敛很慢，为什么？

2.11 用牛顿下山法解 $f(x)=x^2-2=0$ 时，取 $x_0=0.5$，按牛顿迭代公式 $x_{n+1}=\dfrac{1}{2}\left(x_n+\dfrac{2}{x_n}\right)$ 计算出 $x_1=2.25$，此时下山条件不满足，问下山因子 $\lambda$ 为何值时，下山条件能满足？

2.12 证明 $1-x-\sin x=0$ 在[0,1]内有一个根，使用对分法求误差不大于 $0.5\times10^{-4}$ 的根要对分多少次？

2.13 用三阶迭代公式求解方程 $x=e^{-x}$。

# 第三章 解线性方程组的直接法

线性方程组是下列形式的多元一次方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \qquad \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n \end{cases}$$

或简记为

$$AX = B$$

其中

$$A = \begin{bmatrix} a_{11} & a_{12}\cdots a_{1n} \\ a_{21} & a_{22}\cdots a_{2n} \\ \cdots & \\ a_{n1} & a_{n2}\cdots a_{nn} \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}, \quad B = \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_n \end{bmatrix}$$

这里假定系数矩阵 $A$ 的行列式 $|A| \neq 0$,线性方程组的解唯一。

线性方程组的数值解法可以分为直接法和迭代法两类。所谓直接法,就是通过有限步精确运算即能求得线性方程组准确解的方法。这种方法尽管在理论上是完善的算法,但因实际计算时总是有舍入误差存在,所以用直接法所得的结果仍然是准确解的近似值。本章介绍计算机上常用而有效的直接法:消元法和主元素法。

## §1 消元法

### 1.1 方法的一般描述

消元法是求解线性方程组的常用方法,为了分析消元法,今以三个变量的线性方程组为例说明之,其运算规律可以推广到 $n$ 个变量的线性方程组上去。设有线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3 \end{cases} \tag{3.1}$$

消元法的基本思想是通过组合方程的方法实现逐步消元,达到将原方程组化为三角形方程组的目的,然后用回代法解此三角形方程组即可获得原方程组的解。

#### 1.1.1 消元计算过程

对于方程组(3.1),首先取用其第一个方程(3.1),分别与式(3.1)的其余两个方程进行组合,消去它们方程中的 $x_1$ 所在项。为了便于消元,在消元前,可选用一个常数 $l_{11}$ 遍除式$(3.1)_1$ 得

$(3.1)_1/l_{11}$:

$$u_{11}x_1 + u_{12}x_2 + u_{13}x_3 = z_1 \tag{3.2}$$

其中

$$u_{11} = \frac{a_{11}}{l_{11}}, \quad u_{12} = \frac{a_{12}}{l_{11}}, \quad u_{13} = \frac{a_{13}}{l_{11}}, \quad z_1 = \frac{b_1}{l_{11}}$$

然后引入以下两个乘数

$$l_{21}=\frac{a_{21}}{u_{11}},\quad l_{31}=\frac{a_{31}}{u_{11}}$$

按以下组合方式消去式$(3.1)_2$与式$(3.1)_3$中的$x_1$项：

$(3.1)_2-l_{21}\cdot(3.2)$：　$0+(a_{22}-l_{21}u_{12})x_2+(a_{23}-l_{21}u_{13})x_3=b_2-l_{21}z_1$

$(3.1)_3-l_{31}\cdot(3.2)$：　$0+(a_{32}-l_{31}u_{12})x_2+(a_{33}-l_{31}u_{13})x_3=b_3-l_{31}z_1$

简记为

$$\begin{cases}a_{22}^{(1)}x_2+a_{23}^{(1)}x_3=b_2^{(1)}\\a_{32}^{(1)}x_2+a_{33}^{(1)}x_3=b_3^{(1)}\end{cases}\tag{3.3}$$

其中

$$\begin{cases}a_{ij}^{(1)}=a_{ij}^{(0)}-l_{i1}u_{1j}\\b_i^{(1)}=b_i^{(0)}-l_{i1}z_1\\a_{ij}^{(0)}=a_{ij},\quad b_i^{(0)}=b_i\end{cases}\tag{3.4}$$

对于方程组(3.3)，类似地用其第一个方程$(3.3)_1$与(3.3)的其余方程进行组合，消去$x_2$项。消元前，选用一个常数$l_{22}$遍除式$(3.3)_1$得$(3.3)_1/l_{22}$：

$$u_{22}x_2+u_{23}x_3=z_2\tag{3.5}$$

其中

$$u_{22}=\frac{a_{22}^{(1)}}{l_{22}}=\frac{a_{22}-l_{21}u_{12}}{l_{22}},\quad u_{23}=\frac{a_{23}^{(1)}}{l_{22}}=\frac{a_{23}-l_{21}u_{13}}{l_{22}}$$

$$z_2=\frac{b_2^{(1)}}{l_{22}}=\frac{b_2-l_{21}z_1}{l_{22}}$$

再引入乘数

$$l_{32}=\frac{a_{32}^{(1)}}{u_{22}}=\frac{a_{32}-l_{31}u_{12}}{u_{22}}$$

按以下组合方式消去式$(3.3)_2$中的$x_2$项：

$$(3.3)_2-l_{32}\cdot(3.5):\quad 0+(a_{33}^{(1)}-l_{32}u_{23})x_3=b_3^{(1)}-l_{32}z_2$$

简记为

$$a_{33}^{(2)}x_3=b_3^{(2)}\tag{3.6}$$

其中

$$\begin{cases}a_{33}^{(2)}=a_{33}^{(1)}-l_{32}u_{23}=a_{33}-l_{31}u_{13}-l_{32}u_{23}\\b_3^{(2)}=b_3^{(1)}-l_{32}z_2=b_3-l_{31}z_1-l_{32}z_2\end{cases}\tag{3.7}$$

同法对式(3.6)遍除$l_{33}$得

$$u_{33}x_3=z_3\tag{3.8}$$

其中

$$\begin{cases}u_{33}=\dfrac{a_{33}^{(2)}}{l_{33}}=\dfrac{a_{33}-l_{31}u_{13}-l_{32}u_{23}}{l_{33}}\\z_3=\dfrac{b_3^{(2)}}{l_{33}}=\dfrac{b_3-l_{31}z_1-l_{32}z_2}{l_{33}}\end{cases}\tag{3.9}$$

以上的计算过程称为消元过程。消元过程结束就可得到下列三角形线性方程组

$$\begin{cases}u_{11}x_1+u_{12}x_2+u_{13}x_3=z_1\\\qquad\quad\ u_{22}x_2+u_{23}x_3=z_2\\\qquad\qquad\qquad u_{33}x_3=z_3\end{cases}\tag{3.10}$$

简记为

$$\boldsymbol{UX}=\boldsymbol{Z}\tag{3.11}$$

其中

$$\boldsymbol{U}=\begin{bmatrix}u_{11}&u_{12}&u_{13}\\0&u_{22}&u_{23}\\0&0&u_{33}\end{bmatrix},\quad \boldsymbol{X}=\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix},\quad \boldsymbol{Z}=\begin{bmatrix}z_1\\z_2\\z_3\end{bmatrix}$$

而

$$\begin{cases} z_1 = \dfrac{b_1}{l_{11}} \\ z_2 = \dfrac{b_2 - l_{21} z_1}{l_{22}} \\ z_3 = \dfrac{b_3 - l_{31} z_1 - l_{32} z_2}{l_{33}} \end{cases}$$

它可以改写为下列线性方程组：

$$\begin{cases} l_{11} z_1 & = b_1 \\ l_{21} z_1 + l_{22} z_2 & = b_2 \\ l_{31} z_1 + l_{32} z_2 + l_{33} z_3 & = b_3 \end{cases} \tag{3.12}$$

简记为

$$\boldsymbol{LZ} = \boldsymbol{B} \tag{3.13}$$

其中

$$\boldsymbol{L} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix}$$

由 $\boldsymbol{LZ}=\boldsymbol{B}$，$\boldsymbol{Z}=\boldsymbol{UX}$ 得 $\boldsymbol{LUX}=\boldsymbol{B}$，与 $\boldsymbol{AX}=\boldsymbol{B}$ 比较知

$$\boldsymbol{A} = \boldsymbol{LU} \tag{3.14}$$

可见，消元过程实质上就是将原线性方程组分解为两个三角形线性方程组 $\boldsymbol{LZ}=\boldsymbol{B}$ 和 $\boldsymbol{UX}=\boldsymbol{Z}$ 的计算过程。两个线性方程组的系数矩阵均为三角形矩阵，$\boldsymbol{L}$ 为左下三角形矩阵，$\boldsymbol{U}$ 为右上三角形矩阵，在数值上具有关系式(3.14)，因此消元的计算过程也可以说是将 $\boldsymbol{A}$ 分解为 $\boldsymbol{L}$ 与 $\boldsymbol{U}$ 的计算过程。

### 1.1.2　消元过程的计算公式

根据消元过程的计算顺序和运算规律，可以列出它们的计算公式如表 3.1 所示。表中各数值元素的计算规则归纳如下。

① $l_{ij}$，$u_{ij}$ 的分子部分由 $a_{ij}$ 减去若干项内积构成；而 $z_i$ 的分子部分由 $b_i$ 减去若干项内积构成。为说明起见，以 $l_{ij}$ 为例说明其分子部分中若干内积项的计算规则，它是由 $l_{ij}$ 所在行(即 $i$ 行)左边第一个元素 $l_{i1}$(记为左$_1$)与其所在列(即 $j$ 列)的第一个元素 $u_{1j}$(记为顶$_1$)，然后左$_2$(即 $l_{i2}$)、顶$_2$(即 $u_{2j}$)，左$_3$、顶$_3$…双双能成对的内积所构成。至于 $u_{ij}$ 与 $z_i$ 分子部分中所含有的内积项的计算规则与 $l_{ij}$ 完全相同。

② 表中元素按第一行第一列，第二行第二列…的顺序进行计算。同行中 $u_{ij}$ 的分母均为 $l_{ii}$；同列中 $l_{ij}$ 的分母均为 $u_{jj}$。

上述计算规则对 $n$ 阶线性方程组同样适用，其一般计算公式可表为

$$l_{ij} = \frac{a_{ij} - \sum\limits_{k=1}^{i-1} l_{ik} u_{kj}}{u_{jj}} \qquad (i \geqslant j) \tag{3.15}$$

$$u_{ij} = \frac{a_{ij} - \sum\limits_{k=1}^{i-1} l_{ik} u_{kj}}{l_{ii}} \qquad (i \leqslant j) \tag{3.16}$$

$$z_i = \frac{b_i - \sum\limits_{k=1}^{i-1} l_{ik} z_k}{l_{ii}} \qquad (i = 1,2,\cdots,n) \tag{3.17}$$

表 3.1

<table>
<tr><td>$u_{11}=a_{11}/l_{11}$</td><td rowspan="2">$u_{12}=\frac{a_{12}}{l_{11}}$</td><td rowspan="2">$u_{13}=\frac{a_{13}}{l_{11}}$</td><td rowspan="2">$z_1=\frac{b_1}{l_{11}}$</td></tr>
<tr><td>$l_{11}$(择定)</td></tr>
<tr><td rowspan="2">$l_{21}=\frac{a_{21}}{u_{11}}$</td><td>$u_{22}=(a_{22}-l_{21}u_{12})/l_{22}$</td><td rowspan="2">$u_{23}=\frac{a_{23}-l_{21}u_{13}}{l_{22}}$</td><td rowspan="2">$z_2=\frac{b_2-l_{21}z_1}{l_{22}}$</td></tr>
<tr><td>$l_{22}$(择定)</td></tr>
<tr><td rowspan="2">$l_{31}=\frac{a_{31}}{u_{11}}$</td><td rowspan="2">$l_{32}=\frac{a_{32}-l_{31}u_{12}}{u_{22}}$</td><td>$u_{33}=(a_{33}-l_{31}u_{13}-l_{32}u_{23})/l_{33}$</td><td rowspan="2">$z_3=\frac{b_3-l_{31}z_1-l_{32}z_2}{l_{33}}$</td></tr>
<tr><td>$l_{33}$(择定)</td></tr>
</table>

### 1.1.3　回代计算过程

在 $u_{ij}$ 及 $z_i(i=1,2,\cdots,n)$ 已知的基础上,可建立求解 $x_1,x_2,\cdots,x_n$ 的三角形线性方程组

$$\begin{cases} u_{11}x_1+u_{12}x_2+\cdots+u_{1n}x_n=z_1 \\ \qquad u_{22}x_n+\cdots+u_{2n}x_n=z_2 \\ \qquad\quad \ddots \qquad \cdots \\ \qquad\qquad u_{nn}x_n=z_n \end{cases} \tag{3.18}$$

按由下而上的方程次序解出 $x_n,x_{n-1},\cdots,x_2,x_1$ 如下

$$\begin{cases} x_n=\dfrac{z_n}{u_{nn}} \\ x_{n-1}=\dfrac{z_{n-1}-u_{(n-1)n}x_n}{u_{(n-1)(n-1)}} \\ \cdots \\ x_1=\dfrac{z_1-u_{12}x_2-u_{13}x_3-\cdots-u_{1n}x_n}{u_{11}} \end{cases} \tag{3.19}$$

或

$$x_i=\frac{z_i-\sum\limits_{k=i+1}^{n}u_{ik}x_k}{u_{ii}} \quad (i=n,n-1,\cdots,2,1) \tag{3.20}$$

以上的计算过程称为回代过程。

综上可知,消元法由消元过程和回代过程组成。在消元过程中,$l_{ii}(i=1,2,\cdots,n)$ 的数值是可以任意取定的,对于 $l_{ii}$ 值的不同取法就产生了不同的消元法,它们选值的不同会影响到计算量及舍入误差的大小。常用的有以下三种消元法。

## 1.2　高斯消元法

在这种消元法中,取 $l_{ii}=1(i=1,2,\cdots,n)$,相应的计算公式为

$$
\begin{cases}
l_{ij} = \dfrac{a_{ij} - \sum\limits_{k=1}^{j-1} l_{ik} u_{kj}}{u_{jj}} & (i > j) \\
u_{ij} = a_{ij} - \sum\limits_{k=1}^{i-1} l_{ik} u_{kj} & (i \leqslant j) \\
z_i = b_i - \sum\limits_{k=1}^{i-1} l_{ik} z_k & (i = 1,2,\cdots,n) \\
x_i = \dfrac{z_i - \sum\limits_{k=i+1}^{n} u_{ik} x_k}{u_{ii}} & (i = n,n-1,\cdots,2,1)
\end{cases} \tag{3.21}
$$

**例 3.1** 用高斯消元法解下列线性方程组

$$
\begin{cases}
-23x_1 + 11x_2 + x_3 = 0 \\
11x_1 - 3x_2 - 2x_3 = 3 \\
x_1 - 2x_2 + 2x_3 = -1
\end{cases} \tag{3.22}
$$

**解** 按高斯消元法的计算公式(3.21)得 $l_{ij}$、$u_{ij}$、$z_i$ 的数值列于表 3.2，由此便可建立以下三角形线性方程组

$$
\begin{cases}
-23x_1 + 11x_2 + x_3 = 0 \\
2.260\,86x_2 - 1.521\,74x_3 = 3 \\
1.019\,24x_3 = 1.019\,21
\end{cases}
$$

逐次回代解得

$$x_3 = \frac{1.019\,21}{1.019\,24} = 0.999\,97$$

$$x_2 = \frac{3-(-1.521\,74)\times 0.999\,97}{2.260\,86} = 1.999\,99$$

$$x_1 = \frac{0-11\times 1.999\,99-1\times 0.999\,97}{-23} = 0.999\,99$$

**表 3.2**

| | | | |
|---|---|---|---|
| $u_{11}=-23$<br>$l_{11}=1$ | $u_{12}=11$ | $u_{13}=1$ | $z_1=0$ |
| $l_{21}=\dfrac{11}{-23}$ $=-0.478\,26$ | $u_{22}=-3-(-0.478\,26)\times 11$ $=2.260\,86$<br>$l_{22}=1$ | $u_{23}=-2-(-0.478\,26)\times 1$ $=-1.521\,74$ | $z_2=3-(-0.478\,26)\times 0$ $=3$ |
| $l_{31}=\dfrac{1}{-23}$ $=-0.043\,48$ | $l_{32}=\dfrac{-2-(-0.043\,48)\times 11}{2.260\,86}$ $=-0.673\,07$ | $u_{33}=2-(-0.043\,48)\times 1-(-0.673\,07)\times(-1.521\,74)$ $=1.019\,24$<br>$l_{33}=1$ | $z_3=-1-(-0.043\,48)\times 0-(-0.673\,07)\times 3$ $=1.019\,21$ |

### 1.3　克劳特消元法

在这种消元法中，取 $l_{11}=a_{11}^{(0)}=a_{11}$，$l_{22}=a_{22}^{(1)}$，$l_{33}=a_{33}^{(2)}$，…，$l_{nn}=a_{nn}^{(n-1)}$，相应地有 $u_{ii}=1(i=1,2,\cdots,n)$。计算公式如下

$$\begin{cases} l_{ij}=a_{ij}-\sum\limits_{k=1}^{j-1}l_{ik}u_{kj} & (i\geqslant j) \\ u_{ij}=\dfrac{a_{ij}-\sum\limits_{k=1}^{i-1}l_{ik}u_{kj}}{l_{ii}} & (i<j) \\ z_i=\dfrac{b_i-\sum\limits_{k=1}^{i-1}l_{ik}z_k}{l_{ii}} & (i=1,2,\cdots,n) \\ x_i=z_i-\sum\limits_{k=i+1}^{n}u_{ik}x_k & (i=n,n-1,\cdots,2,1) \end{cases} \tag{3.23}$$

**例 3.2**　用克劳特消元法解下列线性方程组

$$\begin{cases} 2x_1+3x_2+4x_3=6 \\ 3x_1+5x_2+2x_3=5 \\ 4x_1+3x_2+30x_3=32 \end{cases}$$

**解**　按克劳特消元法的计算公式(3.23)建立 $l_{ij}$、$u_{ij}$、$z_i$ 的数值表 3.3。按表值列出以下三角形线性方程组

$$\begin{cases} x_1+1.5x_2+2x_3=3 \\ \qquad\quad x_2-8x_3=-8 \\ \qquad\qquad\quad x_3=2 \end{cases}$$

解得 $x_3=2$，$x_2=-8+8\times2=8$，$x_1=3-1.5\times8-2\times2=-13$

**表 3.3**

<table>
<tr><td>$u_{11}=1$</td><td rowspan="2">$u_{12}=\dfrac{3}{2}=1.5$</td><td rowspan="2">$u_{13}=\dfrac{4}{2}=2$</td><td rowspan="2">$z_1=\dfrac{6}{2}=3$</td></tr>
<tr><td>$l_{11}=2$</td></tr>
<tr><td rowspan="2">$l_{21}=3$</td><td>$u_{22}=1$</td><td rowspan="2">$u_{23}=\dfrac{2-3\times2}{0.5}=-8$</td><td rowspan="2">$z_2=\dfrac{5-3\times3}{0.5}=-8$</td></tr>
<tr><td>$l_{22}=5-3\times1.5=0.5$</td></tr>
<tr><td rowspan="2">$l_{31}=4$</td><td rowspan="2">$l_{32}=3-4\times1.5=-3$</td><td>$u_{33}=1$</td><td rowspan="2">$z_3=\dfrac{32-4\times3-(-3)\times(-8)}{-2}=2$</td></tr>
<tr><td>$l_{33}=30-4\times2-(-3)\times(-8)=-2$</td></tr>
</table>

### 1.4　平方根法

本法是针对系数矩阵为对称的线性方程组设计的，在这种消元法中，取 $l_{ii}=u_{ii}(i=1,2,\cdots,n)$，由于线性方程组具有对称的系数矩阵，因此 $a_{ij}=a_{ji}$。在这种情况下，$l_{ij}$ 与 $u_{ij}$ 的数值如表 3.4 所示。

**表 3.4**

| $l_{11}=u_{11}$ | $u_{12}=\frac{a_{12}}{l_{11}}=l_{21}$ | $u_{13}=\frac{a_{13}}{l_{11}}=l_{31}$ | $u_{14}=\frac{a_{14}}{l_{11}}=l_{41}$ | … |
|---|---|---|---|---|
| $l_{21}=\frac{a_{21}}{u_{11}}$ | $l_{22}=u_{22}$ | $u_{23}=\frac{a_{23}-l_{21}l_{31}}{l_{22}}=l_{32}$ | $u_{24}=\frac{a_{24}-l_{21}l_{41}}{l_{22}}=l_{42}$ | … |
| $l_{31}=\frac{a_{31}}{u_{11}}$ | $l_{32}=\frac{a_{32}-l_{31}l_{21}}{u_{22}}$ | $l_{33}=u_{33}$ | $u_{34}=\frac{a_{34}-l_{31}l_{41}-l_{32}l_{42}}{l_{33}}=l_{43}$ | … |
| $l_{41}=\frac{a_{41}}{u_{11}}$ | $l_{42}=\frac{a_{42}-l_{41}l_{21}}{u_{22}}$ | $l_{43}=\frac{a_{43}-l_{41}l_{31}-l_{42}l_{32}}{u_{33}}$ | $l_{44}=u_{44}$ | … |
| … | … | … | … | ⋱ |

由表 3.4 可见，$l_{ij}$ 与 $u_{ij}$ 相对于对角线是对称分布的，由此得 $u_{ij}=l_{ji}$。又因

$$l_{ii}=\frac{a_{ii}-\sum\limits_{k=1}^{i-1}l_{ik}u_{ki}}{u_{ii}}=\frac{a_{ii}-\sum\limits_{k=1}^{i-1}l_{ik}l_{ik}}{l_{ii}}\qquad(i=j)$$

所以

$$l_{ii}=\sqrt{a_{ii}-\sum_{k=1}^{i-1}l_{ik}^2}\tag{3.24}$$

由于 $l_{ii}$ 的数值是通过根式来计算的，因此称这种消元法为平方根法。平方根法的计算公式如下

$$\begin{cases}l_{ij}=\dfrac{a_{ij}-\sum\limits_{k=1}^{j-1}l_{ik}u_{kj}}{l_{ii}}=\dfrac{a_{ij}-\sum\limits_{k=1}^{j-1}l_{ik}l_{jk}}{l_{ii}} & (i>j)\\ l_{ii}=\sqrt{a_{ii}-\sum\limits_{k=1}^{i-1}l_{ik}^2} & (i=j)\\ z_i=\dfrac{b_i-\sum\limits_{k=1}^{i-1}l_{ik}z_k}{l_{ii}} & (i=1,2,\cdots,n)\\ x_i=\dfrac{z_i-\sum\limits_{k=i+1}^{n}u_{ik}x_k}{u_{ii}}=\dfrac{z_i-\sum\limits_{k=i+1}^{n}l_{ki}x_k}{l_{ii}} & (i=n,n-1,\cdots,2,1)\end{cases}\tag{3.25}$$

**例 3.3**　用平方根法解线性方程组

$$\begin{cases}x_1+0.42x_2+0.54x_3=0.3\\ 0.42x_1+x_2+0.32x_3=0.5\\ 0.54x_1+0.32x_2+x_3=0.7\end{cases}$$

**解**　按平方根法的计算公式(3.25)建立 $l_{ij}$、$u_{ij}$、$z_i$ 的数值表 3.5。按表值列出以下三角形线性方程组

$$\begin{cases}x_1+0.42x_2+0.54x_3=10.3\\ \quad 0.907\,52x_2+0.102\,70x_3=0.412\,11\\ \quad\quad 0.835\,37x_3=0.593\,36\end{cases}$$

解得 $x_3=0.710\ 30, x_2=0.373\ 72, x_1=-0.240\ 52$。

表 3.5

| | | | |
|---|---|---|---|
| $l_{11}=u_{11}=\sqrt{1}=1$ | $u_{12}=l_{21}=0.42$ | $u_{13}=l_{31}=0.54$ | $z_1=\frac{0.3}{1}=0.3$ |
| $l_{21}=\frac{0.42}{1}=0.42$ | $l_{22}=u_{22}=\sqrt{1-0.42^2}$ $=0.907\ 52$ | $u_{23}=l_{32}=0.102\ 70$ | $z_2=\frac{0.5-0.42\times0.3}{0.907\ 52}$ $=0.412\ 11$ |
| $l_{31}=\frac{0.54}{1}=0.54$ | $l_{32}=\frac{0.32-0.54\times0.42}{0.907\ 52}$ $=0.102\ 70$ | $l_{33}=\sqrt{1-0.54^2-0.102\ 70^2}$ $=0.835\ 37=u_{33}$ | $z_3=\frac{0.7-0.54\times0.3-0.102\ 70\times0.412\ 11}{0.835\ 37}$ $=0.593\ 36$ |

在采用平方根法的求解过程中，当根式内的数值 $l_{ii}^2>0$ 时，则 $l_{ij}$ 全为实数；但当根式内的某些数值 $l_{ii}^2<0$ 时，$l_{ii}$ 为虚数，则以下的运算要按复数运算规则进行。

## 1.5　追赶法

在实际问题中，如样条插值及常微分方程边值问题的数值解中，都会遇到求解三对角线形的线性方程组：

$$\begin{cases} b_1x_1+c_1x_2 & =d_1 \\ a_2x_1+b_2x_2+c_2x_3 & =d_2 \\ \quad a_3x_2+b_3x_3+c_3x_4 & =d_3 \\ \quad\ddots\quad\ddots\quad\ddots & \vdots \\ \quad a_{n-1}x_{n-2}+b_{n-1}x_{n-1}+c_{n-1}x_n & =d_{n-1} \\ \quad a_nx_{n-1}+b_nx_n & =d_n \end{cases} \tag{3.26}$$

这个线性方程组可应用克劳特消元法求解之。

### 1.5.1　$l_{ij}, u_{ij}$ 的数值计算公式

将克劳特消元法的计算公式应用于方程组(3.26)可得

$$\begin{matrix} l_{11}=b_1 \quad u_{12}=\frac{c_1}{l_{11}} & 0 & 0 & \cdots & 0 \\ l_{21}=a_2 \quad l_{22}=b_2-a_2u_{12} & u_{23}=\frac{c_2}{l_{22}} & 0 & & \vdots \\ \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots \quad l_{(n-1)(n-2)}=a_{n-1} & & l_{(n-1)(n-1)}=b_{n-1}-a_{n-1}u_{(n-2)(n-1)} & & u_{(n-1)n}=\frac{c_{n-1}}{l_{(n-1)(n-1)}} \\ 0 \quad \cdots \quad 0 & & l_{n(n-1)}=a_n & & l_{nn}=b_n-a_nu_{(n-1)n} \end{matrix}$$

这里
$$\begin{cases} l_{i(i-1)} = a_i \quad (i = 2,3,\cdots,n) \\ l_{ii} = b_i - a_i u_{(i-1)i} \quad (i = 1,2,\cdots,n), a_1 = u_{01} = 0 \\ u_{i(i+1)} = \dfrac{c_i}{l_{ii}} \quad (i = 1,2,\cdots,n-1) \end{cases} \tag{3.27}$$

### 1.5.2 $z_i$ 的数值计算公式

根据上述 $l_{ij}$ 的数值可建立如下关系式

$$\begin{cases} l_{11} z_1 = d_1 \\ a_2 z_1 + l_{22} z_2 = d_2 \\ a_3 z_2 + l_{33} z_3 = d_3 \\ \cdots \\ a_{n-1} z_{n-2} + l_{(n-1)(n-1)} z_{n-1} = d_{n-1} \\ a_n z_{n-1} + l_{nn} z_n = d_n \end{cases} \tag{3.28}$$

解得
$$\begin{cases} z_1 = d_1 / l_{11} \\ z_2 = (d_2 - a_2 z_1)/l_{22} \\ z_3 = (d_3 - a_3 z_2)/l_{33} \\ \cdots \\ z_{n-1} = (d_{n-1} - a_{n-1} z_{n-2})/l_{(n-1)(n-1)} \\ z_n = (d_n - a_n z_{n-1})/l_{nn} \end{cases} \tag{3.29}$$

或
$$z_i = \frac{d_i - a_i z_{i-1}}{l_{ii}} \quad (i = 1,2,\cdots,n), z_0 = a_1 = 0 \tag{3.30}$$

### 1.5.3 $x_i$ 的数值计算公式

根据上述 $u_{ij}$ 的数值可建立如下关系式

$$\begin{cases} x_1 + u_{12} x_2 = z_1 \\ x_2 + u_{23} x_3 = z_2 \\ \cdots \\ x_{n-1} + u_{(n-1)n} x_n = z_{n-1} \\ x_n = z_n \end{cases} \tag{3.31}$$

解得
$$\begin{cases} x_1 = z_1 - u_{12} x_2 \\ x_2 = z_2 - u_{23} x_3 \\ \cdots \\ x_{n-1} = z_{n-1} - u_{(n-1)n} x_n \\ x_n = z_n \end{cases} \tag{3.32}$$

或
$$x_i = z_i - u_{i(i+1)} x_{i+1} \quad (i = n, n-1, \cdots, 2, 1) \tag{3.33}$$

以上求解三对角线形线性方程组的解法称为追赶法。其中的消元过程称为追过程；而回代过程称为赶过程。

**例 3.4** 用追赶法解下列线性方程组

$$\begin{cases} -2x_1 + x_2 = -2 \\ x_1 - 2x_2 + x_3 = 1 \\ x_2 - 2x_3 = -4 \end{cases}$$

**解**　按式(3.27)计算得数值表3.6。

**表3.6**

<table>
<tr><td>$u_{11}=1$</td><td rowspan="2">$u_{12}=\frac{1}{-2}=-0.5$</td><td rowspan="2"></td></tr>
<tr><td>$l_{11}=-2$</td></tr>
<tr><td rowspan="2">$l_{21}=1$</td><td>$u_{22}=1$</td><td rowspan="2">$u_{23}=\frac{1}{-1.5}=-0.666\ 67$</td></tr>
<tr><td>$l_{22}=-2-1\times(-0.5)=-1.5$</td></tr>
<tr><td rowspan="2"></td><td rowspan="2">$l_{32}=1$</td><td>$u_{33}=1$</td></tr>
<tr><td>$l_{33}=-2-1\times(-0.666\ 67)=1.333\ 33$</td></tr>
</table>

按式(3.28)建立如下方程组

$$\begin{cases} -2z_1 & =-2 \\ z_1-\ 1.5z_2 & =1 \\ z_2-1.333\ 33z_3 & =-4 \end{cases}$$

解得 $z_1=1, z_2=0, z_3=3.000\ 01$

再按式(3.31)建立方程组

$$\begin{cases} x_1-0.5x_2 & =1 \\ x_2-0.666\ 67x_3 & =0 \\ x_3 & =3.000\ 01 \end{cases}$$

解得 $x_3=3.000\ 01, x_2=2.000\ 02, x_3=2.000\ 01$

## 1.6　消元法的应用条件

为保证消元法的运算过程顺利进行，由消元法的计算公式可见，必须要求计算公式中的分母不等于零，即要求 $l_{ii}\neq 0, u_{ii}\neq 0(i=1,2,\cdots,n)$。否则上述情况出现时，在电子数字计算机上因分母为零而导致计算过程中断。为此，我们来推证 $l_{ii}\neq 0, u_{ii}\neq 0$ 所应满足的条件。

**定理3.1**　若 $\boldsymbol{A}$ 的各阶主子式均不为零，即

$$|\boldsymbol{A}_1|=|a_{11}|\neq 0, |\boldsymbol{A}_2|=\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}\neq 0, |\boldsymbol{A}_3|=\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}\neq 0,\cdots, |\boldsymbol{A}_n|=|\boldsymbol{A}|\neq 0 \tag{3.34}$$

时，则 $l_{ii}\neq 0, u_{ii}\neq 0(i=1,2,\cdots,n)$。

证：由前知，消元法的计算过程实质上就是将 $\boldsymbol{A}$ 转化为 $\boldsymbol{LU}$ 乘积的过程。若 $\boldsymbol{A}$ 可转化为 $\boldsymbol{LU}$，则其各阶主子矩阵 $\boldsymbol{A}_i(i=1,2,\cdots,n)$ 均能转化成相应阶的积 $\boldsymbol{L}_i\boldsymbol{U}_i$，即有

$$\boldsymbol{A}_1=\boldsymbol{L}_1\boldsymbol{U}_1=[l_{11}][u_{11}]$$

$$\boldsymbol{A}_2=\boldsymbol{L}_2\boldsymbol{U}_2=\begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix}\begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix}$$

$$\cdots$$

$$A_n = L_n U_n = LU = \begin{bmatrix} l_{11} & & & 0 \\ l_{21} & l_{22} & & \\ \vdots & & \ddots & \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ & u_{22} & \cdots & u_{2n} \\ 0 & & \ddots & \vdots \\ & & & u_{nn} \end{bmatrix} \tag{3.35}$$

于是有

$$\begin{cases} |A_1| = l_{11}u_{11} \\ |A_2| = l_{11}l_{22}u_{11}u_{22} = |A_1| l_{22}u_{22} \\ |A_3| = l_{11}l_{22}l_{33}u_{11}u_{22}u_{33} = |A_2| l_{33}u_{33} \\ \cdots \\ |A_n| = |A_{n-1}| l_{nn}u_{nn} \end{cases} \tag{3.36}$$

据定理 3.1 条件，由 $|A_1| \neq 0$，可推知 $l_{11} \neq 0, u_{11} \neq 0$；由 $|A_1| \neq 0$、$|A_2| \neq 0$，可以推知 $l_{22} \neq 0$，$u_{22} \neq 0$；由 $|A_2| \neq 0$、$|A_3 \neq 0$，可以推知 $l_{33} \neq 0, u_{33} \neq 0$…由 $|A_{n-1}| \neq 0$、$|A_n| \neq 0$，可以推知 $l_{nn} \neq 0, u_{nn} \neq 0$。(证毕)

**定理 3.2** 若 $A$ 为实对称正定矩阵，则 $l_{ii} \neq 0, u_{ii} \neq 0 (i=1,2,\cdots,n)$。

证：因实对称正定矩阵 $A$ 为正定的必要且充分条件是 $A$ 的各阶主子式都大于零，即

$$|A_1| > 0, |A_2| > 0, \cdots, |A_n| > 0 \tag{3.37}$$

显然满足定理 3.1 的条件(3.34)，即 $|A_i| \neq 0 (i=1,2,\cdots,n)$，因此定理 3.2 得证。

当线性方程组的系数矩阵 $A$ 为对称正定且采用平方根法求解时，则必有 $l_{ii}^2 > 0 (i=1,2,\cdots,n)$。这是因为

$$0 < |A_1| = [l_{11}][l_{11}] = l_{11}^2$$

$$0 < |A_2| = \left| \begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} \\ 0 & l_{22} \end{bmatrix} \right| = \begin{vmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{vmatrix} \begin{vmatrix} l_{11} & l_{21} \\ 0 & l_{22} \end{vmatrix} = l_{11}^2 l_{22}^2 \text{，所以 } l_{22}^2 > 0$$

$$\cdots$$

$$0 < |A_n| = l_{11}^2 l_{22}^2 \cdots l_{(n-1)(n-1)}^2 l_{nn}^2 \text{，故 } l_{nn}^2 > 0$$

因此在求解过程中不可能出现复数运算的现象。

**定理 3.3** 若 $A$ 为严格对角占优矩阵，则 $l_{ii} \neq 0, u_{ii} \neq 0 (i=1,2,\cdots,n)$。

证：所谓严格对角占优矩阵，指的是其对角线上元素的绝对值大于同行上其余元素绝对值之和的矩阵，即满足以下不等式

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \qquad (i = 1,2,\cdots,n) \tag{3.38}$$

的矩阵。在这种情况下，$A$ 的各阶主子矩阵 $A_1$、$A_2$、…、$A_n$ 亦均是严格对角占优矩阵，据阿达玛定理知，各阶主子式 $|A_i| \neq 0 (i=1,2,\cdots,n)$，据定理 3.1，证得 $l_{ii} \neq 0, u_{ii} \neq 0 \ (i=1,2,\cdots,n)$。

下面附证阿达玛定理。

**阿达玛定理** $r$ 阶主子式 $|A_r| \neq 0$ 的一个充分条件是下述严格对角占优条件

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{r} |a_{ij}| \qquad (i = 1,2,\cdots,r) \tag{3.39}$$

成立。

证：用反证法证明之，假设 $|A_r| = 0$，则线性方程组 $A_r X = 0$ 有非零解 $\alpha_1, \alpha_2, \cdots, \alpha_r$。设

$$|\alpha_k| = \max(|\alpha_1|, |\alpha_2|, \cdots, |\alpha_r|)$$

则 $\alpha_k$ 满足下式

$$a_{kk}\alpha_k = -\sum_{\substack{j=1\\j\neq k}}^{r} a_{kj}\alpha_j$$

$$|a_{kk}| \leqslant \sum_{\substack{j=1\\j\neq k}}^{r} |a_{kj}| \cdot \left|\frac{\alpha_j}{\alpha_k}\right| \leqslant \sum_{\substack{j=1\\j\neq k}}^{r} |a_{kj}| \tag{3.40}$$

式(3.40)与式(3.39)条件矛盾,故假设 $|\boldsymbol{A}_r|=0$ 不能成立,定理得证。

## §2　选主元的高斯消元法

前述的消元法中,未知量是按出现在方程中的自然顺序消去的,用来消去其他方程式中未知量的方程亦是按顺序取定的,所以又叫顺序消元法。实际计算已发现有以下缺点。以高斯消元法为例,消元过程中可能出现 $u_{kk}=a_{kk}^{(k-1)}=0$ 的情况,则在计算乘数 $l_{ik}=a_{ik}^{(k-1)}/u_{kk}(i=k+1,k+2,\cdots,n)$ 时发生计算中断现象。即使 $u_{kk}\neq 0$,但当其值甚小时,就会使乘数 $|l_{ik}|\gg 1$ 而导致舍入误差的严重扩大。因此在消元法的应用条件中要求 $A$ 的各阶主子式不为 0,以保证 $u_{ii}\neq 0(i=1,2,\cdots,n)$。而线性方程组的解存在唯一只需 $|\boldsymbol{A}|\neq 0$ 已够,上述要求完全是方法本身在使用上的限制所造成的。下面叙述的主元素法就是针对以上计算问题而提出的解决方法,最常用和最有效的主元素法有列主元素法和全主元素法。

### 2.1　列主元素法

在消元过程中,为排除出现 $u_{kk}=0$ 而导致的障碍,可以简单地通过交换方程次序的办法来解决。即将 $a_{ik}^{(k-1)}\neq 0$ 所在的方程与 $u_{kk}$ 所在的方程(即第 $k$ 个方程)进行交换。经这样交换后,便可获得新的不为 0 的 $u_{kk}$ 值,以下便可继续实现消元计算。在消元计算中为使舍入误差减小,就应使乘数 $|l_{ik}|$ 值尽量地小,这就是说,应在 $k$ 列元素中选取绝对值最大的元素作为新的 $u_{kk}$ 值,这个绝对值最大的元素称为主元素。从以上分析,不难得出以下方程交换的原则:应在 $k$ 列中将主元素所在的方程与第 $k$ 个方程进行交换,使主元素位于第 $k$ 个对角线位置上。我们把这种使用主元素的消元法称为列主元素法。下面举例说明之。

**例 3.5**　用列主元素法解下列线性方程组

$$\begin{cases} 10x_1 - 19x_2 - 2x_3 = 3 \\ \boxed{-20}\ x_1 + 40x_2 + x_3 = 4 \\ x_1 + 4x_2 + 5x_3 = 5 \end{cases} \tag{3.41}$$

**解**　首先从方程组(3.41)的第一列系数 10,−20,1 中选出绝对值最大的元素 −20 作为该列的主元素,交换第一、二两个方程的位置使该主元素位于对角线的第一个位置上,得到

$$\begin{cases} \boxed{-20}\ x_1 + 40x_2 + x_3 = 4 \\ 10x_1 - 19x_2 - 2x_3 = 3 \\ x_1 + 4x_2 + 5x_3 = 5 \end{cases} \tag{3.42}$$

计算乘数

$$l_{21} = \frac{10}{-20} = -0.5,\quad l_{31} = \frac{1}{-20} = -0.05$$

保持$(3.42)_1$不变，消元得

$$\begin{cases}(3.42)_2 - l_{21}(3.42)_1: & 0 + x_2 - 1.5x_3 = 5 \\ (3.42)_3 - l_{31}(3.42)_1: & 0 + \boxed{6}\ x_2 + 5.05x_3 = 5.2\end{cases} \tag{3.43}$$

从方程组(3.43)的第二列系数1,6中选出绝对值最大的元素6作为该列的主元素，交换第二、三两个方程的位置使该主元素位于对角线的第二个位置上得

$$\begin{cases}\boxed{6}\ x_2 + 5.05x_3 = 5.2 \\ x_2 - 1.5x_3 = 5\end{cases} \tag{3.44}$$

计算乘数

$$l_{32} = \frac{1}{6} = 0.166\ 67$$

保持$(3.44)_1$不变，消元得

$$(3.44)_2 - l_{32}(3.44)_1: \quad 0 - \boxed{2.341\ 68}\ x_3 = 4.133\ 32 \tag{3.45}$$

最后，在第三列系数中选主元素，它就是$-2.341\ 68$，这时方程中只含有一个变量$x_3$，消元过程结束。联立主元素所在的方程得

$$\begin{cases}-20x_1 + 40x_2 + x_3 = 4 \\ 6x_2 + 5.05x_3 = 5.2 \\ -2.341\ 68x_3 = 4.133\ 32\end{cases} \tag{3.46}$$

逐次回代求解(3.46)得

$$x_3 = -1.765\ 11, \quad x_2 = 2.352\ 30, \quad x_1 = 4.416\ 34$$

### 2.2 全主元素法

如果不是逐次按列选主元素，而是在全体待选的系数中选取主元素，则得全主元素法。其求解过程以下例说明之。

**例3.6** 用全主元素法解例3.5。

**解** 首先在线性方程组

$$\begin{cases}10x_1 - 19x_2 - 2x_3 = 3 \\ -20x_1 + \boxed{40}\ x_2 + x_3 = 4 \\ x_1 + 4x_2 + 5x_3 = 5\end{cases}$$

的所有系数中取绝对值最大的元素40为主元素，并交换第一、二方程和交换第一、二列使该主元素位于对角线的第一个位置上得

$$\begin{cases}\boxed{40}\ x_2 - 20x_1 + x_3 = 4 \\ -19x_2 + 10x_1 - 2x_3 = 3 \\ 4x_2 + x_1 + 5x_3 = 5\end{cases} \tag{3.47}$$

计算乘数

$$l_{21} = \frac{-19}{40} = -0.475, \quad l_{31} = \frac{4}{40} = 0.1$$

保持$(3.47)_1$不变，消元得

$$\begin{cases}(3.47)_2 - l_{21}(3.47)_1: & 0 + 0.5x_1 - 1.525x_3 = 4.9 \\ (3.74)_3 - l_{31}(3.47)_1: & 0 + 3x_1 + \boxed{4.9}\ x_3 = 4.6\end{cases} \tag{3.48}$$

从方程组(3.48)的所有系数中选取绝对值最大的元素 4.9，交换第二、三两个方程和交换第二、三列使该主元素位于对角线的第二个位置上得

$$\begin{cases}\boxed{4.9}\ x_3 + 3x_1 = 4.6 \\ -1.525x_3 + 0.5x_1 = 4.9\end{cases} \tag{3.49}$$

计算乘数

$$l_{32} = \frac{-1.525}{4.9} = -0.311\,22$$

保持$(3.49)_1$不变，消元得

$$(3.49)_2 - l_{32}(3.49)_1: \quad \boxed{1.433\,66}\ x_1 = 6.331\,61 \tag{3.50}$$

最后取(3.50)中的 1.433 66 为主元素，消元过程结束。联立主元素所在的方程得

$$\begin{cases}40x_2 - 20x_1 + x_3 = 4 \\ 4.9x_3 + 3x_1 = 4.6 \\ 1.433\,66x_1 = 6.331\,6\end{cases} \tag{3.51}$$

逐次回代求解式(3.51)得

$$x_1 = 4.416\,40, \quad x_3 = -1.765\,14, \quad x_2 = 2.352\,33。$$

采用主元素法不仅在消元过程中可以减小舍入误差；而且在回代过程中，由于采用数值较大的主元素作分母，同样可以减小除法运算的误差。因此主元素法具有良好的数值稳定性，其中全主元素法的精度优于列主元素法，因在主元素法每一步均要选取主元素，这会增加工作量。但在列主元素法中，因选取主元素的范围有限且对换方程的次序并不改变方程的同解性，与一般的高斯消元法比较，其增加的运算量不大。与列主元素法相比，全主元素法选取主元素的范围较大，除方程间作对换外，还要进行列交换，列交换后，未知量的次序亦作了对换，这就需要记录下列交换的序号，以便在计算过程结束后恢复原来未知量的序号，其增加的工作量和程序设计的难度更大于列主元素法，这就是一般多采用列主元素法的重要原因。

# §3 关于结果精度的检验

由于实际问题中所提供的数据(系数矩阵和右端项的元素)一般含有观测误差；有的数学问题中的系数矩阵和右端项的元素是前面计算的结果，也会引入误差；最后将数据输入计算机中进行数制转换也会引入舍入误差。原始数据的这些误差均会使线性方程组的精确解发生变化。在直接法的计算模型中，由于使用的方法是精确的，因此不存在方法误差。而计算模型数值解的舍入误差则由两类运算误差构成。一类舍入误差由计算公式中分子部分的内积型 $(\sum a_i b_i)$ 运算生成；另一类舍入误差由除法运算生成。它们在计算过程中经传递和累积形成结果的舍入误差。目前已有对舍入误差的一些估计公式，一般都偏大，不切实用，实际上不会去应用它。应该指出，要完全弄清楚它并不是一件容易的事情。下面仅对结果的精度检验进行一些粗略的分析与讨论。

## 3.1 残差法

我们可把具有误差数据的线性方程组视为参数模型，在用直接法求得上述线性方程组的

近似解 $\tilde{X}$ 后，那么这个近似解与参数模型精确解之间究竟相差多少？这个问题很难定量地回答，因为并不知道上述精确解的数值是多大。因此只能通过一些间接的方式来估计。最简单的估计办法是把近似解 $\tilde{X}$ 代入原来方程组去求出所谓"残差"(残余或余量)$\boldsymbol{r}$

$$\boldsymbol{r} = \boldsymbol{B} - \boldsymbol{A}\tilde{\boldsymbol{X}} \tag{3.52}$$

如果 $\boldsymbol{r}$ 的每个分量 $r_i$ 都是小量，那么，一般就认为近似解是相当准确的，否则认为是不准确的。这种方法简单，运算量少，对大多数实际问题也还是很可靠的。其缺点是从残差的大小无法定量地确定近似解究竟有几位是准确的，所能得出的只是近似解准确与否的一个粗略概念。此外，这样的概念对于线性方程组在病态情况下是不可靠的。例如，我们考虑下列方程组

$$\begin{bmatrix} 0.216\,1 & 0.144\,1 \\ 1.296\,9 & 0.864\,8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.144\,0 \\ 0.864\,2 \end{bmatrix} \tag{3.53}$$

如果以 $\tilde{x}_1=0.991,\tilde{x}_2=-0.487\,0$ 代入(3.53)中，将得到残差

$$\boldsymbol{r} = \begin{bmatrix} -0.000\,000\,01 \\ 0.000\,000\,01 \end{bmatrix}$$

显见，按照上述残差的大小可以认为在取小数后四位数字情况下近似解已足够精确了。然而，这一方程组的精确解却是 $\alpha_1=2,\alpha_2=-2$。这就说明，尽管残差已经很小，近似解的精度还可能很差。如果把(3.53)中的右端项稍微变化一点，改为

$$\tilde{\boldsymbol{B}} = \begin{bmatrix} 0.144\,000\,01 \\ 0.864\,199\,99 \end{bmatrix}$$

则精确解由 $\alpha_1=2,\alpha_2=-2$ 变为前述的 $\tilde{x}_1=0.991,\tilde{x}_2=-0.487\,0$。这里，右端项的微小变化引起了参数模型精确解的巨大变化，因此线性方程组(3.53)具有"病态"性。按残差法来判断近似解精确度的办法对于病态线性方程组来说一般是不可靠的。在线性方程组为病态情况下，由于数据的舍入误差对参数模型精确解具有较大的影响，一般可采用高精度计算来削弱其病态的程度。

## 3.2 类比法

另一种衡量近似解精度的办法是用类比法判定它的有效数位数。方法是先任取一个已知向量 $\boldsymbol{Z}$，用较多的位数计算出向量 $\bar{\boldsymbol{B}}=\boldsymbol{A}\boldsymbol{Z}$，再以 $\bar{\boldsymbol{B}}$ 为右端项，建立线性方程组

$$\boldsymbol{A}\bar{\boldsymbol{X}} = \bar{\boldsymbol{B}} \tag{3.54}$$

按某种直接法求解(3.54)的数值解，如果计算过程无舍入误差，$\bar{\boldsymbol{X}}$ 应与 $Z$ 相等。但因求解过程中有舍入误差的积累，使 $\bar{\boldsymbol{X}}$ 与 $Z$ 间有差异，这种差异可用 $\bar{\boldsymbol{X}}$ 与 $Z$ 相一致的有效数字位数来度量，这样既可把 $\bar{\boldsymbol{X}}$ 与 $Z$ 各分量中相符合的最少位数 $N$ 取定为该直接法数值解可能达到的有效数字位数。

对于 $\boldsymbol{AX}=\boldsymbol{B}$ 按上述相同的直接法求解时，一般认为舍入误差对结果的影响是相同的，因此所得的数值解的有效数位可按 $N$ 来取定。这种方法比较可靠，能获得近似解有几位有效数字的一个数量概念，缺点是计算量大。

# 习 题 三

3.1　用高斯消元法解下列方程组

$$\begin{cases}3x_1+2x_2+5x_3=6\\-x_1+4x_2+3x_3=5\\x_1-x_2+3x_3=1\end{cases}$$

3.2　用克劳特消元法求解下列方程组

$$\begin{cases}3x_1-x_2+4x_3=7\\-x_1+2x_2-2x_3=-1\\2x_1-3x_2-2x_3=0\end{cases}$$

3.3　用平方根法解下列方程组

$$\begin{cases}3x_1-x_2+2x_3=7\\-x_1+2x_2-2x_3=-1\\2x_1-2x_2+4x_3=0\end{cases}$$

3.4　用列主元素法和全主元素法解方程组

$$\begin{cases}3x_1-x_2+4x_3=3\\-x_1+2x_2-2x_3=2\\2x_1-3x_2-2x_3=-5\end{cases}$$

3.5　用追赶法解下列方程组

$$\begin{cases}2x_1-x_2=0\\-x_1+2x_2-x_3=1\\-x_2+2x_3-x_4=0\\-x_3-2x_4=2.5\end{cases}$$

3.6　用平方根法解

$$\begin{cases}16x_1+4x_2+8x_3=-4\\4x_1+5x_2-4x_3=3\\8x_1-4x_2+22x_3=10\end{cases}$$

3.7　用追赶法解方程组 $\boldsymbol{AX}=\boldsymbol{B}$，其中

$$A=\begin{bmatrix}2&-1&&&&&\\-1&2&-1&&&0&\\&-1&2&-1&&&\\&&-1&2&-1&&\\&&&-1&2&-1&\\&&0&&-1&2&-1\\&&&&&-1&2\end{bmatrix},\quad B=\begin{bmatrix}1\\0\\0\\0\\0\\0\\0\end{bmatrix}$$

# 第四章　解线性方程组的迭代法

线性方程组是联立方程组的一种特殊类型的方程组，联立方程组迭代解法的基本思想和方法完全适用于线性方程组的情况。在下面的算法分析与讨论中，经常要遇到向量范数、矩阵范数以及序列极限等概念。为此，首先介绍这方面的一些基本知识。

## §1　向量范数、矩阵范数、谱半径及有关性质

向量范数是用来度量向量长度的，它可以看成是解析几何中二、三维向量长度概念的推广。

**定义 4.1**　对任一向量 $\boldsymbol{X}\in\mathbf{R}^n$，按照一定规则确定一个实数与它对应，该实数记为 $\|\boldsymbol{X}\|$，若 $\|\boldsymbol{X}\|$ 满足下面三个性质：

① $\|\boldsymbol{X}\|\geqslant 0$；$\|\boldsymbol{X}\|=0$ 当且仅当 $\boldsymbol{X}=0$；

② 对任意实数 $\alpha$，$\|\alpha\boldsymbol{X}\|=|\alpha|\,\|\boldsymbol{X}\|$；

③ 对任意向量 $\boldsymbol{X},\boldsymbol{Y}\in\mathbf{R}^n$，$\|\boldsymbol{X}+\boldsymbol{Y}\|\leqslant\|\boldsymbol{X}\|+\|\boldsymbol{Y}\|$。则称该实数 $\|\boldsymbol{X}\|$ 为向量 $\boldsymbol{X}$ 的范数。

在 $\mathbf{R}^n$ 中，常用的几种范数有：

$$\|\boldsymbol{X}\|_1=|x_1|+|x_2|+\cdots+|x_n|=\sum_{i=1}^{n}|x_i| \tag{4.1}$$

$$\|\boldsymbol{X}\|_2=\sqrt{x_1^2+x_2^2+\cdots+x_n^2}=(\sum_{i=1}^{n}x_i^2)^{\frac{1}{2}} \tag{4.2}$$

$$\|\boldsymbol{X}\|_\infty=\max\{|x_1|,|x_2|,\cdots,|x_n|\}=\max_{1\leqslant i\leqslant n}\{|x_i|\} \tag{4.3}$$

式中，$x_1,x_2,\cdots,x_n$ 分别是 $\boldsymbol{X}$ 的 $n$ 个分量。以上定义的范数分别称为 1—范数，2—范数和 $\infty$—范数。可以验证它们都是满足范数性质的，其中 $\|\boldsymbol{X}\|_2$ 是由内积导出的向量范数。这些范数都是 $p$ 范数

$$\|\boldsymbol{X}\|_p=\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}} \tag{4.4}$$

的特例。当不需要指明哪一种向量范数时，就用记号 $\|\cdot\|$ 泛指任何一种向量范数。

有了向量的范数，就可以用它来衡量向量的大小和表示向量的误差。设 $\boldsymbol{\alpha}$ 为 $\boldsymbol{AX}=\boldsymbol{B}$ 的精确解，$\boldsymbol{X}$ 为其近似解，则其绝对误差可表示成 $\|\boldsymbol{X}-\boldsymbol{\alpha}\|$，其相对误差可表示成 $\|\boldsymbol{X}-\boldsymbol{\alpha}\|/\|\boldsymbol{\alpha}\|$ 或 $\|\boldsymbol{X}-\boldsymbol{\alpha}\|/\|\boldsymbol{X}\|$。

从向量范数出发，还可以定义矩阵的范数。矩阵范数是用于表示矩阵“大小”的量，类似于向量范数，可以定义 $n$ 阶方阵 $\boldsymbol{A}$ 的范数。

**定义 4.2**　设 $\boldsymbol{A}$ 为 $n$ 阶方阵，若对应的非负实数 $\|\boldsymbol{A}\|$ 满足：

① $\|\boldsymbol{A}\|\geqslant 0$；$\|\boldsymbol{A}\|=0$ 当且仅当 $\boldsymbol{A}=0$ 时；

② 对任意实数 $\alpha$，$\|\alpha\boldsymbol{A}\|=|\alpha|\,\|\boldsymbol{A}\|$；

③ $\|\boldsymbol{A}+\boldsymbol{B}\|\leqslant\|\boldsymbol{A}\|+\|\boldsymbol{B}\|$；

④ $\|AB\| \leqslant \|A\|\|B\|$。

其中 $B$ 也是 $n$ 阶方阵，则称 $\|A\|$ 为矩阵 $A$ 的范数。

以上前三条性质是与向量范数类似的，第四个性质则是矩阵乘法特点所要求的。

在矩阵计算中，矩阵与向量的乘积经常出现，因此需要把向量范数和矩阵范数联系起来考虑。设 $\mathbf{R}^n$ 中规定的向量范数为 $\|X\|_\alpha$，在 $\mathbf{R}^{n\times n}$ 中规定的矩阵范数为 $\|A\|_\beta$，要求向量范数与矩阵范数满足以下不等式

$$\|AX\|_\alpha \leqslant \|A\|_\beta \|X\|_\alpha \tag{4.5}$$

当以上不等式成立时，便称矩阵范数 $\|A\|_\beta$ 和向量范数 $\|X\|_\alpha$ 相容。

当定义一种矩阵范数时，应当使它能与某种向量范数相容。在同一个问题中要同时使用矩阵范数和向量范数时，这两种范数应当是相容的。现在给出一种定义矩阵范数的方法。

**定理 4.1**　设在 $\mathbf{R}^n$ 中给定了一种向量范数，对任一 $n$ 阶方阵 $A$，令

$$\|A\| = \max_{\|X\|=1} \|AX\| \tag{4.6}$$

则由式(4.6)所定义 $\|\cdot\|$ 是一种矩阵范数，并且它与所给定的向量范数相容。(证明略)

称式(4.6)所定义的矩阵范数为从属于给定向量范数的矩阵范数。这种矩阵范数实际上就是把矩阵看成是 $\mathbf{R}^n$ 上线性变换的算子范数，所以称为矩阵的算子范数或由向量范数导出的矩阵范数。由式(4.6)可看出，任何一个算子范数，当 $A$ 为单位矩阵时，必有 $\|I\|=1$，这是算子范数的必要条件。对于给定的向量 1—范数、2—范数及∞—范数，可以证明从属于它的矩阵范数分别为

$$\|A\|_1 = \max_{1\leqslant j\leqslant n} \sum_{i=1}^{n} |a_{ij}| \tag{4.7}$$

$$\|A\|_2 = \sqrt{\lambda_{\max}(A'A)} \tag{4.8}$$

$$\|A\|_\infty = \max_{1\leqslant i\leqslant n} \sum_{j=1}^{n} |a_{ij}| \tag{4.9}$$

式中，$\lambda_{\max}(A'A)$ 表示矩阵 $A'A$ 的最大特征值。请读者回忆，当 $A^{-1}$ 存在时，矩阵 $A'A$ 是正定的对称矩阵，其特征值全为正；当 $A^{-1}$ 不存在时，$A'A$ 是半正定矩阵，其特征值非负。而 $\|A\|_\infty$ 是 $A$ 的行向量中 1—范数的最大值，简称行范数；$\|A\|_1$ 则是 $A$ 的列向量中 1—范数的最大值，简称列范数。

除以上三种常用的矩阵范数外，还有一种常用的矩阵范数，就是

$$\|A\|_F = \sqrt{\sum_{i,j=1}^{n} a_{ij}^2} \tag{4.10}$$

称 $\|A\|_F$ 为 $A$ 的 $F$—范数(Frobenius 范数)，因为 $\|I\|_F=\sqrt{n}$，所以它不是算子范数，但也满足矩阵范数定义的四个条件，可以证明它与向量范数中的 2—范数相容，即

$$\|AX\|_2 \leqslant \|A\|_F \|X\|_2 \tag{4.11}$$

综上可知，矩阵的从属范数必与给定的向量范数相容，但是矩阵范数与向量范数相容，却未必有从属关系。

有了范数的概念，就可以来叙述收敛的问题。

**定义 4.3**　对于 $\mathbf{R}^n$ 中的向量序列 $\{X^{(k)}\}$，如果

$$\lim_{k\to\infty} \|X^{(k)} - X\| = 0 \tag{4.12}$$

则称向量序列$\{X^{(k)}\}$收敛于$\mathbf{R}^n$中的向量$X$。

**定义 4.4** 对于$n$阶方阵序列$\{A^{(k)}\}$，如果

$$\lim_{k\to\infty}\|A^{(k)}-A\|=0 \tag{4.13}$$

则称矩阵序列$\{A^{(k)}\}$收敛于$n$阶方阵$A$。

式(4.12)有时亦表为

$$\lim_{k\to\infty}X^{(k)}=X \tag{4.14}$$

同样，式(4.13)有时亦表为

$$\lim_{k\to\infty}A^{(k)}=A \tag{4.15}$$

从上面定义可以直接推出下面定理。

**定理 4.2** $\mathbf{R}^n$中的向量序列$\{X^{(k)}\}$收敛于$\mathbf{R}^n$中的向量$X$的必要充分条件是

$$\lim_{k\to\infty}x_j^{(k)}=x_j \qquad (j=1,2,\cdots,n) \tag{4.16}$$

式中，$x_j^{(k)}$和$x_j$分别表示$X^{(k)}$和$X$中的第$j$个分量。定理说明，$\mathbf{R}^n$空间中向量序列的收敛可以归结为各坐标分量的收敛。

**定理 4.3** $n$阶方阵序列$\{A^{(k)}\}$收敛于$n$阶方阵$A$的必要充分条件是

$$\lim_{k\to\infty}a_{ij}^{(k)}=a_{ij} \qquad (i,j=1,2,\cdots,n) \tag{4.17}$$

这个定理告诉我们，矩阵序列的收敛也可归结为对应元素序列的收敛。

**定义 4.5** 设$n$阶方阵$A$的特征值为$\lambda_i(i=1,2,\cdots,n)$，则称

$$\rho(A)=\max_{1\leqslant i\leqslant n}|\lambda_i| \tag{4.18}$$

为矩阵$A$的谱半径。

对于任何一种具有相容向量范数的矩阵范数$\|A\|$成立不等式

$$\rho(A)\leqslant\|A\| \tag{4.19}$$

这是因为矩阵$A$的任一特征值$\lambda_i$与其对应的特征向量$X_i$间有关系式

$$AX_i=\lambda_iX_i$$

两端取范数，再利用性质(4.5)得

$$|\lambda_i|\,\|X_i\|\leqslant\|A\|\,\|X_i\|$$

由于$X_i\neq0$，则$\|X_i\|\neq0$，所以$|\lambda_i|\leqslant\|A\|$，故

$$\rho(A)\leqslant\|A\|$$

下面只考虑具有相容向量范数的矩阵范数，上式说明，矩阵$A$的谱半径不超过它的任何一种范数，即$\|A\|$是$A$的特征值的上界。对于矩阵$A$的2—范数有以下定理。

**定理 4.4** 如果$A\in\mathbf{R}^{n\times n}$，则

① $\|A\|_2=\sqrt{\lambda_{\max}(A'A)}=\sqrt{\rho(A'A)}$ (4.20)

② 若$A$为对称矩阵，则

$$\|A\|_2=\rho(A)$$

**证** 显然$A'A$是对称矩阵，对任何$X\in\mathbf{R}^n$，$\|AX\|_2^2=(AX)'(AX)=X'(A'A)X$

而$\|AX\|_2^2\geqslant0$，所以$X'(A'A)X\geqslant0$，即$A'A$是对称半正定矩阵。令$A'A$的特征值为

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n\geqslant0$$

则它们均大于等于零。记对应的标准正交特征向量为$e_1,e_2,\cdots,e_n$并取为$\mathbf{R}^n$的一个基底，设

$A'Ae_i=\lambda_i e_i$，则对任何$\|X\|_2=1$的$X$可表为

$$X = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n$$

计算

$$\begin{aligned}\|AX\|_2^2 &= (AX)'(AX) = X'(A'A)X \\ &= (x_1 e_1 + \cdots + x_n e_n)'(A'A)(x_1 e_1 + \cdots + x_n e_n) \\ &= (x_1 e_1 + \cdots + x_n e_n)'(\lambda_1 x_1 e_1 + \cdots + \lambda_n x_n e_n) \\ &= (x_1 e'_1 + \cdots + x_n e'_n)(\lambda_1 x_1 e_1 + \cdots + \lambda_n x_n e_n) \\ &= \lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2 \\ &\leqslant \lambda_1 (x_1^2 + x_2^2 + \cdots + x_n^2) \qquad (\lambda_1 = \max_i \lambda_i) \\ &= \lambda_1\end{aligned}$$

另一方面，取$X=e_1$，则$\|AX\|_2^2=\lambda_1$，即上界可达。从而证得

$$\|A\|_2^2 = \max_{\|X\|=1} \|AX\|_2 = \lambda_1 = \lambda_{\max}(A'A) = \rho(A'A)$$

特别当$A$为对称矩阵时，有$A'=A, A'A=A^2$，记$A$的特征值为$\mu_1, \mu_2, \cdots, \mu_n$，且$|\mu_1| \geqslant |\mu_2| \geqslant \cdots \geqslant |\mu_n|$，则有$\lambda_i=\mu_i^2$（因为$A'A=A^2$），就有

$$\|A\|_2 = \sqrt{\lambda_{\max}(A'A)} = \sqrt{\lambda_{\max}(A^2)} = \sqrt{\max_{1\leqslant i\leqslant n} \mu_i^2} = \max_{1\leqslant i\leqslant n} |\mu_i| = \rho(A) = |\mu_1|$$

（证毕）

由于2—范数具有关系式(4.20)，所以$\|A\|_2$被称为谱范数。

**定理4.5**　设$A$是任意$n$阶方阵，由$A$的各次幂所组成的矩阵序列

$$I,\ A,\ A^2,\ \cdots,\ A^k, \cdots \tag{4.21}$$

收敛于零，则$\lim\limits_{k\to\infty} A^k=0$的必要充分条件是

$$\rho(A) < 1 \tag{4.22}$$

证明略。

本章叙述最常用的几种迭代法，包括简单迭代法、赛德尔迭代法、松弛迭代法以及迭代法的收敛性与精度控制的问题。使用迭代法求解线性方程组，具有计算简单、编制程序容易、存储量较小、舍入误差积累小（只需计算最终迭代那一次的舍入误差）等优点，较适合于高阶线性方程组的求解。

# §2　简单迭代法

## 2.1　迭代公式

对于线性方程组$AX=B$，首先应将其改写为$X=\Phi(X)$的形式，其方法是多种多样的。下面介绍两种常用的迭代格式。

### 2.1.1　迭代格式1

将$AX=B$改写为$0=B-AX$，两边加上$X$后得

$$X = [I - A]X + B = CX + B \tag{4.23}$$

或写成

$$x_i = \sum_{j=1}^{n} c_{ij} x_j + b_i \qquad (i = 1, 2, \cdots, n) \tag{4.24}$$

其中 $c_{ij}=-a_{ij}(i\neq j)$，$c_{ii}=1-a_{ii}(i=j)$。相应的迭代公式为

$$X^{(k+1)}=CX^{(k)}+B \tag{4.25}$$

或

$$x_i^{(k+1)}=\sum_{j=1}^{n}c_{ij}x_j^{(k)}+b_i \qquad (i=1,2,\cdots,n) \tag{4.26}$$

式中 $C=I-A$ 称为迭代矩阵，它等于

$$C=\begin{bmatrix}1-a_{11} & -a_{12} & -a_{13} & \cdots & -a_{1n}\\ -a_{21} & 1-a_{22} & -a_{23} & \cdots & -a_{2n}\\ \cdots & \cdots & \cdots & & \cdots\\ -a_{n1} & -a_{n2} & -a_{n3} & \cdots & 1-a_{nn}\end{bmatrix} \tag{4.27}$$

### 2.1.2 迭代格式 2

若 $a_{ii}\neq 0(i=1,2,\cdots,n)$，可按照方程的顺序依次地直接解出 $x_1,x_2,\cdots,x_n$ 得

$$\begin{aligned}x_i&=\sum_{j=1}^{i-1}\left(-\frac{a_{ij}}{a_{ii}}\right)x_j+\sum_{j=i+1}^{n}\left(-\frac{a_{ij}}{a_{ii}}\right)x_j+\frac{b_i}{a_{ii}}\\ &=\sum_{j=1}^{n}g_{ij}x_j+f_i \qquad (i=1,2,\cdots,n)\end{aligned} \tag{4.28}$$

其中

$$g_{ij}=-\frac{a_{ij}}{a_{ii}}(i\neq j),\qquad g_{ii}=0(i=j),\qquad f_i=\frac{b_i}{a_{ii}}\text{。若令}$$

$$G=\begin{bmatrix}0 & g_{12} & g_{13} & \cdots & g_{1n}\\ g_{21} & 0 & g_{23} & \cdots & g_{2n}\\ \vdots & \vdots & \vdots & & \vdots\\ g_{n1} & g_{n2} & g_{n3} & \cdots & 0\end{bmatrix},\quad D=\begin{bmatrix}a_{11} & & & 0\\ & a_{22} & & \\ 0 & & \ddots & \\ & & & a_{nn}\end{bmatrix},\quad F=\begin{bmatrix}f_1\\ f_2\\ \vdots\\ f_n\end{bmatrix}$$

则式(4.28)可用矩阵表示为

$$X=GX+F \tag{4.29}$$

容易看出

$$G=D^{-1}(D-A)=I-D^{-1}A,\qquad F=D^{-1}B \tag{4.30}$$

相应的迭代公式为

$$X^{(k+1)}=GX^{(k)}+F \tag{4.31}$$

或

$$x_i^{(k+1)}=\sum_{j=1}^{n}g_{ij}x_j^{(k)}+f_i \qquad (i=1,2,\cdots,n) \tag{4.32}$$

这种迭代法称为雅可比迭代法。

对于上述建立的迭代公式，任取一组初值 $X^{(0)}=(x_1^{(0)},x_2^{(0)},\cdots,x_n^{(0)})'$ 作为根 $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_n)'$ 的零次近似值，按迭代公式进行迭代计算。如果迭代序列 $X^{(k+1)}$ 有极限存在，则此极限即为线性方程组的根。称这种解法为简单迭代法。

## 2.2 简单迭代法的收敛条件

实际使用的迭代法应该是收敛的迭代法，下面要给出收敛的判别条件。为了叙述的方便，把 $AX=B$ 改写后的等价方程组统一表为

$$X = MX + N \tag{4.33}$$

或

$$x_i = \sum_{j=1}^{n} m_{ij}x_j + n_i = \varphi_i(x_1, x_2, \cdots, x_n) \qquad (i = 1, 2, \cdots, n) \tag{4.34}$$

其中

$$M = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \cdots & \cdots & & \cdots \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{bmatrix}, \qquad N = \begin{bmatrix} n_1 \\ n_2 \\ \cdots \\ n_n \end{bmatrix}$$

下面给出保证收敛的有关定理和条件。

**定理 4.6**　对任何初始向量 $X^{(0)}$ 和常数项 $N$，由迭代公式

$$X^{(k+1)} = MX^{(k)} + N, \qquad (k = 0, 1, 2, \cdots) \tag{4.35}$$

产生的向量序列 $\{X^{(k)}\}$ 收敛的必要充分条件是

$$\rho(M) < 1 \tag{4.36}$$

式中，$\rho(M)$ 是矩阵 $M$ 的谱半径。

证：*必要性*　设序列 $\{X^{(k)}\}$ 收敛于 $\alpha$，则有

$$\alpha = M\alpha + N$$

第 $k$ 次迭代的近似值和精确解之差为

$$X^{(k+1)} - \alpha = MX^{(k)} - M\alpha = M(X^{(k)} - \alpha) \tag{4.37}$$

反复使用上式得

$$X^{(k+1)} - \alpha = M(X^{(k)} - \alpha) = M^2(M^{(k-1)} - \alpha) = \cdots = M^{k+1}(X^{(0)} - \alpha) \tag{4.38}$$

对于任意初始误差向量 $(X^{(0)} - \alpha)$，为使

$$\lim_{k\to\infty}(X^{(k+1)} - \alpha) = 0$$

必须

$$\lim_{k\to\infty} M^k = 0$$

由定理 4.5 即知 $\rho(M) < 1$。

*充分性*　若 $\rho(M) < 1$ 满足，则特征值 $|\lambda| < 1$，不发生 $|I - M| = 0$ 的问题，则 $I - M$ 非奇异，从而方程组 $(I - M)X = N$ 有唯一解，设为 $\alpha$，这时式(4.37)仍成立，重复同样的推证知

$$\lim_{k\to\infty} X^{(k+1)} = \alpha$$

即迭代过程收敛。(证毕)

从上述定理看出，迭代的收敛性只与迭代矩阵的谱半径有关，而迭代矩阵是由 $A$ 演变来的，因此迭代是否收敛是与系数矩阵 $A$ 以及演变的方式有关，与常数项和初始向量的选择无关。

定理 4.6 给出了简单迭代法收敛的必要充分条件是迭代矩阵的谱半径小于 1，但在具体问题中，谱半径是很难计算的，用它来判定收敛性是不现实的。下面给出几个容易使用的判断收敛的充分条件及有关的误差估计公式。

由式(4.34)知

$$\left|\frac{\partial \varphi_i}{\partial x_j}\right| = |m_{ij}| \qquad (i, j = 1, 2, \cdots, n) \tag{4.39}$$

应用第二章 §4 联立方程组迭代解法收敛的充分条件，可以得到关于简单迭代法的三个收敛充分条件如下。

**充分条件 1** 若 $\mu=\|M\|_\infty<1$，则对任意初值，简单迭代法收敛。且

$$\|X^{(k)}-\alpha\|_\infty \leqslant \frac{\mu^k}{1-\mu}\|X^{(1)}-X^{(0)}\|_\infty \tag{4.40}$$

**充分条件 2** 若 $\nu=\|M\|_1<1$，则对任意初值，简单迭代法收敛。且

$$\|X^{(k)}-\alpha\|_1 \leqslant \frac{\nu^k}{1-\nu}\|X^{(1)}-X^{(0)}\|_1 \tag{4.41}$$

**充分条件 3** 若 $p=\|M\|_F<1$，则对任意初值，简单迭代法收敛。且

$$\|X^{(k)}-\alpha\|_2 \leqslant \frac{p^k}{1-p}\|X^{(1)}-X^{(0)}\|_2 \tag{4.42}$$

与联立方程组迭代解法局部收敛充分条件不同，在线性方程组的情况下，由式(4.39)可见，$|\partial\varphi_i/\partial x_j|$ 在任意初值下都为常数，因此，建立在此基础上的三个充分条件与初值的选取无关，属大范围收敛充分条件。这三个充分条件可以统一用下面的定理来描述。

**定理 4.7** 若迭代矩阵 $M$ 的范数 $\|M\|=q<1$，则简单迭代法收敛。且迭代序列 $\{X^{(k)}\}$ 的第 $k$ 次近似值与精确解 $\alpha$ 的误差有估计式

$$\|X^{(k)}-\alpha\| \leqslant \frac{q^k}{1-q}\|X^{(1)}-X^{(0)}\| \tag{4.43}$$

为了使误差 $\|X^{(k)}-\alpha\|$ 小于要求的精度 $\varepsilon$，可以利用这一估计式来计算需要的迭代次数，但一般都偏大，实用上常采用下面的定理。

**定理 4.8** 若 $\|M\|<1$，则迭代序列 $\{X^{(k)}\}$ 的第 $k$ 次近似值 $X^{(k)}$ 和精确解 $\alpha$ 的误差有估计式

$$\|X^{(k)}-\alpha\| \leqslant \frac{\|M\|}{1-\|M\|}\|X^{(k)}-X^{(k-1)}\| \tag{4.44}$$

**证** 由于

$$\begin{aligned} X^{(k)}-\alpha &= MX^{(k-1)}+N-M\alpha-N \\ &= MX^{(k-1)}-M\alpha \end{aligned} \tag{4.45}$$

由 $\alpha=M\alpha+N$ 得 $\alpha=(I-M)^{-1}N$，代入上式得

$$\begin{aligned} X^{(k)}-\alpha &= MX^{(k-1)}-M(I-M)^{-1}N \\ &= M(I-M)^{-1}[(I-M)X^{k-1}-N] \\ &= M(I-M)^{-1}[X^{(k-1)}-X^{(k)}] \end{aligned}$$

两边取范数，即得

$$\|X^{(k)}-\alpha\| \leqslant \|M\|\,\|(I-M)^{-1}\|\,\|X^{(k)}-X^{(k-1)}\| \tag{4.46}$$

下面估计 $\|(I-M)^{-1}\|$，因为

$$\begin{aligned} (I-M)(I-M)^{-1} &= I \\ (I-M)^{-1}-M(I-M)^{-1} &= I \\ (I-M)^{-1} &= I+M(I-M)^{-1} \end{aligned} \tag{4.47}$$

两边取范数并利用矩阵范数的性质得

$$\begin{cases} \|(I-M)^{-1}\| \leqslant \|I\|+\|M(I-M)^{-1}\| \leqslant 1+\|M\|\,\|(I-M)^{-1}\| \\ (1-\|M\|)\|(I-M)^{-1}\| \leqslant 1 \end{cases} \tag{4.48}$$

因 $\|M\|<1$，$1-\|M\|>0$，所以上式可化为

$$\|(I-M)^{-1}\| \leqslant \frac{1}{1-\|M\|} \tag{4.49}$$

利用式(4.49)，由式(4.46)就可推得如下结果

$$\|X^{(k)}-\alpha\|\leqslant\frac{\|M\|}{1-\|M\|}\|X^{(k)}-X^{(k-1)}\| \qquad \text{（证毕）}$$

有了上面定理，在实际计算时，若允许误差是 $\varepsilon$，我们只需要求相邻两次迭代向量的差满足关系式

$$\|X^{(k)}-X^{(k-1)}\|\leqslant\varepsilon_1 \tag{4.50}$$

那么迭代即可停止，这里

$$\varepsilon_1\leqslant\frac{1-\|M\|}{\|M\|}\varepsilon \tag{4.51}$$

在简单迭代法中，对于雅可比迭代法尚有其单独的收敛充分条件，我们在下面用定理 4.9 给出。在推证以前，先引进与矩阵有关的两个概念。

**定义 4.6** 若矩阵 $A$ 不能通过行的次序的调换和相应列的次序的调换成为[1]

$$\begin{bmatrix}A_{11} & A_{12}\\ 0 & A_{22}\end{bmatrix} \tag{4.52}$$

式中，$A_{11}$，$A_{22}$ 为方阵。则称 $A$ 为不可约矩阵；否则称为可约矩阵。

当 $A$ 为可约矩阵时，则原来的线性方程组可以分割为阶数较低的两个线性方程组。

**定义 4.7** 若矩阵 $A$ 的对角线元素满足

$$|a_{ii}|\geqslant\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}| \qquad (i=1,2,\cdots,n) \tag{4.53}$$

且至少有一个 $i$ 值，使上式中有严格不等号成立，则称 $A$ 具有对角占优。

**引理** 若 $A$ 不可约，且具有对角占优，则 $A$ 为非奇异矩阵且 $a_{ii}\neq0(i=1,2,\cdots,n)$。(证略)

**定理 4.9** 若系数矩阵 $A$ 不可约且具有对角占优，则雅可比迭代法必定收敛。

**证** 要证明雅可比迭代法收敛，根据定理 4.6，只要证明 $\rho(G)<1$ 即可，$G$ 是雅可比迭代法的迭代矩阵。

用反证法。设矩阵 $G$ 有某个特征值 $\lambda$，其 $|\lambda|\geqslant1$，因 $\lambda$ 是 $G$ 的特征值，所以它必满足特征方程

$$|\lambda I-G|=0$$

由于 $A$ 不可约，且具有对角占优，所以 $a_{ii}\neq0(i=1,2,\cdots,n)$，即有 $|D|\neq0$，因此 $D^{-1}$ 存在。因

$$\begin{aligned}\lambda I-G=\lambda I-(I-D^{-1}A)&=\lambda I-I+D^{-1}A\\ &=D^{-1}(\lambda D+A-D)\end{aligned}$$

两边取行列式得

$$|\lambda I-G|=|D^{-1}|\,|\lambda D+A-D|=0$$

记矩阵 $\bar{G}$ 为

$$\bar{G}=\lambda D+A-D=\begin{bmatrix}\lambda a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & \lambda a_{22} & \cdots & a_{2n}\\ \cdots & \cdots & & \cdots\\ a_{n1} & a_{n2} & \cdots & \lambda a_{nn}\end{bmatrix} \tag{4.54}$$

[1] 相应的含义为，将矩阵第 $i$ 行第 $j$ 行互换后，再接第 $i$ 列第 $j$ 列互换，即线性代数中仅限于互换方式的合同变换。

因$|D^{-1}|\neq 0$,必有

$$|\bar{G}| = |\lambda D + A - D| = 0 \tag{4.55}$$

由于矩阵$\bar{G}$中零元素的位置与矩阵$A$中零元素的位置全同,由$A$的不可约性即可推得矩阵$\bar{G}$的不可约性。

由于$|\lambda|\geqslant 1$及$A$具有对角占优,所以有

$$|\lambda a_{ii}| \geqslant |a_{ii}| \geqslant \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| \qquad (i = 1,2,\cdots,n) \tag{4.56}$$

并且至少有一个$i$使不等号严格成立。(4.56)式表明,矩阵$\bar{G}$也具有对角占优,则据引理知

$$|\bar{G}| = |\lambda D + A - D| \neq 0$$

这与式(4.55)相矛盾,故$G$的特征值的模不能大于等于1。定理得证。

由定理4.9可知,当线性方程组的系数矩阵具有不可约、对角占优时,可采用雅可比迭代格式进行迭代计算,则必收敛。对于$A$为不可约、非对角占优的线性方程组,往往只要适当对调方程的次序,还可兼用组合方程的方法,达到化成不可约、对角占优的等价方程组目的。例如下列线性方程组

$$\begin{cases} 2x_1 + 3x_2 - 4x_3 + x_4 = 3 \\ x_1 - 2x_2 - 5x_3 + x_4 = 2 \\ 5x_1 - 3x_2 + x_3 - 4x_4 = 1 \\ 10x_1 + 2x_2 - x_3 + 2x_4 = -4 \end{cases} \tag{4.57}$$

不具有对角占优的特性,采用交换方程次序与组合方程的方法可获得以下对角占优的线性方程组

$$\begin{cases} 10x_1 + 2x_2 - x_3 + 2x_4 = -4 & (4.57)_4 \\ x_1 + 5x_2 + x_3 = 1 & (4.57)_1 - (4.57)_2 \\ x_1 - 2x_2 - 5x_3 + x_4 = 2 & (4.57)_2 \\ 2x_1 - 5x_2 - x_3 - 9x_4 = 9 & 2(4.57)_3 - (4.57)_4 + (4.57)_1 \end{cases}$$

在组合过程中,为确保等价性,新的线性方程组中应包含原线性方程组的每一个方程至少一次。

在计算机上,当使用雅可比迭代法时,为防止溢出,在编程前应整理公式,使$a_{ii}\neq 0(i=1,2,\cdots,n)$。为使迭代过程收敛快,$a_{ii}$的绝对值应尽可能地大。

**例4.1** 对于下列线性方程组

$$\begin{cases} 10x_1 - x_2 - 2x_3 = 7.2 \\ -x_1 + 10x_2 - 2x_3 = 8.3 \\ -x_1 - x_2 + 5x_3 = 4.2 \end{cases} \tag{4.58}$$

按以下方程组

$$\begin{cases} x_1 = 0.1x_2 + 0.2x_3 + 0.72 \\ x_2 = 0.1x_1 + 0.2x_3 + 0.83 \\ x_3 = 0.2x_1 + 0.2x_2 + 0.84 \end{cases} \tag{4.59}$$

建立迭代公式

$$\begin{cases} x_1^{(k+1)} = 0.1x_2^{(k)} + 0.2x_3^{(k)} + 0.72 \\ x_2^{(k+1)} = 0.1x_1^{(k)} + 0.2x_3^{(k)} + 0.83 \\ x_3^{(k+1)} = 0.2x_1^{(k)} + 0.2x_2^{(k)} + 0.84 \end{cases} \tag{4.60}$$

试分析迭代过程的收敛性。

**解** 在本例中,迭代矩阵为

$$\boldsymbol{M} = \begin{bmatrix} 0 & 0.1 & 0.2 \\ 0.1 & 0 & 0.2 \\ 0.2 & 0.2 & 0 \end{bmatrix}$$

其特征方程为

$$\begin{aligned} |\boldsymbol{M} - \lambda \boldsymbol{I}| &= \left| \begin{bmatrix} 0 & 0.1 & 0.2 \\ 0.1 & 0 & 0.2 \\ 0.2 & 0.2 & 0 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right| \\ &= \begin{vmatrix} -\lambda & 0.1 & 0.2 \\ 0.1 & -\lambda & 0.2 \\ 0.2 & 0.2 & -\lambda \end{vmatrix} = -\lambda^3 + 0.09\lambda + 0.008 \\ &= -(\lambda + 0.1)(\lambda^2 - 0.1\lambda - 0.08) = 0 \end{aligned}$$

解得 $\lambda_1 = -0.1, \lambda_2 = 0.337, \lambda_3 = -0.237$。因 $|\lambda_i| < 1 (i=1,2,3)$,按定理 4.6 判知,上述迭代法是收敛的。

# §3 赛德尔迭代法

## 3.1 迭代公式

对于分解式(4.34),我们可按赛德尔迭代方式(2.168)将它表示为如下的迭代公式:

$$\begin{cases} x_1^{(k+1)} = m_{11}x_1^{(k)} + m_{12}x_2^{(k)} + \cdots + m_{1n}x_n^{(k)} + n_1 \\ x_2^{(k+1)} = m_{21}x_1^{(k+1)} + m_{22}x_2^{(k)} + \cdots + m_{2n}x_n^{(k)} + n_2 \\ x_3^{(k+1)} = m_{31}x_1^{(k+1)} + m_{32}x_2^{(k+1)} + m_{33}x_3^{(k)} + \cdots + m_{3n}x_n^{(k)} + n_3 \\ \cdots \\ x_n^{(k+1)} = m_{n1}x_1^{(k+1)} + m_{n2}x_2^{(k+1)} + \cdots + m_{n(n-1)}x_{n-1}^{(k+1)} + m_{nn}x_n^{(k)} + n_n \end{cases} \tag{4.61}$$

或
$$x_i^{(k+1)} = \sum_{j=1}^{i-1} m_{ij}x_j^{(k+1)} + \sum_{j=i}^{n} m_{ij}x_j^{(k)} + n_i \qquad (i = 1,2,\cdots,n; k = 0,1,2,\cdots) \tag{4.62}$$

亦可用矩阵记为

$$\boldsymbol{X}^{(k+1)} = \boldsymbol{M}_1\boldsymbol{X}^{(k+1)} + \boldsymbol{M}_2\boldsymbol{X}^{(k)} + \boldsymbol{N} \qquad (\boldsymbol{M} = \boldsymbol{M}_1 + \boldsymbol{M}_2). \tag{4.63}$$

其中

$$\boldsymbol{X}^{(k+1)} = \begin{bmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ \cdots \\ x_n^{(k+1)} \end{bmatrix}, \quad \boldsymbol{X}^{(k)} = \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ \cdots \\ x_n^{(k)} \end{bmatrix}, \quad \boldsymbol{M}_1 = \begin{bmatrix} 0 & & & \\ m_{21} & 0 & & 0 \\ m_{31} & m_{32} & 0 & \\ \cdots & & \ddots & \ddots \\ m_{n1} & \cdots & m_{n(n-1)} & 0 \end{bmatrix},$$

$$M_2=\begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ & m_{22} & \cdots & m_{2n} \\ 0 & & \ddots & \cdots \\ & & & m_{nn} \end{bmatrix},\quad N=\begin{bmatrix} n_1 \\ n_2 \\ \cdots \\ n_n \end{bmatrix}$$

如果采用迭代格式 2,则 $AX=B$ 所对应的赛德尔迭代公式为

$$x_i^{(k+1)}=-\frac{1}{a_{ii}}\Big[\sum_{j=1}^{i-1}a_{ij}x_j^{(k+1)}+\sum_{j=i+1}^{n}a_{ij}x_j^{(k)}-b_i\Big]\qquad (i=1,2,\cdots,n;\ k=0,1,2,\cdots)\tag{4.64}$$

其矩阵形式为

$$X^{(k+1)}=-D^{-1}LX^{(k+1)}-D^{-1}UX^{(k)}+D^{-1}B\tag{4.65}$$

其中

$$A=L+D+U$$

$$L=\begin{bmatrix} 0 & & & \\ a_{21} & 0 & & 0 \\ \vdots & & \ddots & \\ a_{n1} & \cdots & a_{n(n-1)} & 0 \end{bmatrix},\quad D=\begin{bmatrix} a_{11} & & 0 \\ & a_{22} & \\ 0 & & \ddots & \\ & & & a_{nn} \end{bmatrix},\quad U=\begin{bmatrix} 0 & a_{12} & \cdots & a_{1n} \\ & 0 & & \vdots \\ 0 & & \ddots & a_{n(n-1)} \\ & & & 0 \end{bmatrix}$$

称式(4.64)的迭代方法为高斯-赛德尔迭代法。

**例 4.2** 应用高斯-赛德尔迭代法解例 4.1。

**解** 在本例中,高斯-赛德尔迭代公式为

$$\begin{cases} x_1^{(k+1)}=0.1(x_2^{(k)}+2x_3^{(k)}+7.2) \\ x_2^{(k+1)}=0.1(x_1^{(k+1)}+2x_3^{(k)}+8.3) \\ x_3^{(k+1)}=0.2(x_1^{(k+1)}+x_2^{(k+1)}+4.2) \end{cases}\tag{4.66}$$

取 $x_1^{(0)}=x_2^{(0)}=x_3^{(0)}=0$,按式(4.66)计算如下

$$\begin{cases} x_1^{(1)}=0.1(0+2\times 0+7.2)=0.72 \\ x_2^{(1)}=0.1(0.72+2\times 0+8.3)=0.902 \\ x_3^{(1)}=0.2(0.72+0.902+4.2)=1.164\ 4 \end{cases}$$

$$\begin{cases} x_1^{(2)}=0.1(0.902+2\times 1.164\ 4+7.2)=1.043\ 1 \\ x_2^{(2)}=0.1(1.043\ 1+2\times 1.164\ 4+8.3)=1.167\ 2 \\ x_3^{(2)}=0.2(1.043\ 1+1.167\ 2+4.2)=1.282\ 1 \end{cases}$$

以下迭代结果为

$$\begin{cases} x_1^{(3)}=1.093\ 1 \\ x_2^{(3)}=1.195\ 7, \\ x_3^{(3)}=1.297\ 8 \end{cases}\begin{cases} x_1^{(4)}=1.099\ 1 \\ x_2^{(4)}=1.199\ 5, \\ x_3^{(4)}=1.299\ 7 \end{cases}\begin{cases} x_1^{(5)}=1.099\ 9 \\ x_2^{(5)}=1.199\ 9, \\ x_3^{(5)}=1.300\ 0 \end{cases}\begin{cases} x_1^{(6)}=1.099\ 9 \\ x_2^{(6)}=1.199\ 9 \\ x_3^{(6)}=1.300\ 0 \end{cases}$$

## 3.2 赛德尔迭代法的收敛条件

### 3.2.1 必要充分条件

我们把式(4.63)改写为

$$(I-M_1)X^{(k+1)}=M_2X^{(k)}+N\tag{4.67}$$

$$X^{(k+1)} = (I-M_1)^{-1}M_2X^k + (I-M_1)^{-1}N \tag{4.68}$$

从上式可见，赛德尔迭代法相当于迭代矩阵为$(I-M_1)^{-1}M_2$ 的简单迭代法。由定理 4.6 知，赛德尔迭代法对于任意初值 $X^{(0)}$ 和常数项 $N$ 都收敛的必要充分条件是迭代矩阵$(I-M_1)^{-1}M_2$ 的谱半径小于 1。

### 3.2.2　充分条件 1

若 $\mu=\|M\|_\infty<1$，则对任意初值，赛德尔迭代法收敛，且

$$\|X^{(k)}-\alpha\|_\infty \leqslant \frac{(\mu^*)^k}{1-\mu^*}\|X^{(1)}-X^{(0)}\|_\infty \tag{4.69}$$

其中
$$\mu^* = \max_i \frac{\mu_i}{1-r_i},\quad r_i = \sum_{j=1}^{i-1}|m_{ij}|,\quad \mu_i = \sum_{j=i}^{n}|m_{ij}|$$

对于 $i=1,2,\cdots,n$，相应的 $r_i$ 与 $\mu_i$ 如图 4.1 所示。

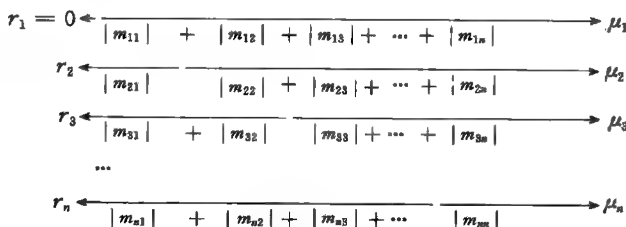



图 4.1

**证**　因

$$x_i^{(k+1)} = \sum_{j=1}^{i-1}m_{ij}x_j^{(k+1)} + \sum_{j=i}^{n}m_{ij}x_j^{(k)} + n_i \tag{4.70}$$

$$\alpha_i = \sum_{j=1}^{n}m_{ij}\alpha_j + n_i \tag{4.71}$$

两式相减得

$$x_i^{(k+1)}-\alpha_i = \sum_{j=1}^{i-1}m_{ij}(x_j^{(k+1)}-\alpha_j) + \sum_{j=i}^{n}m_{ij}(x_j^{(k)}-\alpha_j) \tag{4.72}$$

于是有

$$|x_i^{(k+1)}-\alpha_i| \leqslant \sum_{j=1}^{i-1}|m_{ij}||x_j^{(k+1)}-\alpha_j| + \sum_{j=i}^{n}|m_{ij}||x_j^{(k)}-\alpha_j|,\ (i=1,2,\cdots,n) \tag{4.73}$$

记

$$\|X^{(k+1)}-\alpha\|_\infty = \max_i|x_i^{(k+1)}-\alpha_i|$$

显然
$$|x_i^{(k+1)}-\alpha_i| \leqslant \|X^{(k+1)}-\alpha\|_\infty,\qquad (i=1,2,\cdots,n) \tag{4.74}$$

及
$$|x_i^{(k+1)}-\alpha_i| \leqslant r_i\|X^{(k+1)}-\alpha\|_\infty + \mu_i\|X^{(k)}-\alpha\|_\infty \tag{4.75}$$

在式(4.74)中，设 $i=s$ 时等式成立

$$|x_s^{(k+1)}-\alpha_s| = \max_i|x_i^{(k+1)}-\alpha_i| = \|X^{(k+1)}-\alpha\|_\infty$$

在式(4.75)中令 $i=s$ 得

$$\|X^{(k+1)}-\alpha\|_\infty \leqslant r_s\|X^{(k+1)}-\alpha\|_\infty + \mu_s\|X^{(k)}-\alpha\|_\infty$$

或
$$\|X^{(k+1)}-\alpha\|_\infty \leqslant \frac{\mu_s}{1-r_s}\|X^{(k)}-\alpha\|_\infty \leqslant \mu^*\|X^{(k)}-\alpha\|_\infty \tag{4.76}$$

其中 $$\mu^* = \max_i \frac{\mu_i}{1-r_i}$$

今证 $\mu^*<1$。因 $r_i+\mu_i=\sum_{j=1}^{n}|m_{ij}|\leqslant\mu<1$，所以

$$\mu_i \leqslant \mu - r_i \leqslant \mu - \mu r_i = \mu(1-r_i) \tag{4.77}$$

即得 $$\frac{\mu_i}{1-r_i} \leqslant \mu < 1 \qquad (i=1,2,\cdots,n) \tag{4.78}$$

上式对 $i=1,2,\cdots,n$ 均成立，因此有下式成立

$$\mu^* = \max_i \frac{\mu_i}{1-r_i} \leqslant \mu < 1 \tag{4.79}$$

再由式(4.76)出发，反复使用关系式(4.76)得

$$\|\boldsymbol{X}^{(k+1)}-\boldsymbol{\alpha}\|_\infty \leqslant \mu^*\|\boldsymbol{X}^{(k)}-\boldsymbol{\alpha}\|_\infty \leqslant (\mu^*)^2\|\boldsymbol{X}^{(k-1)}-\boldsymbol{\alpha}\|_\infty \leqslant \cdots \leqslant (\mu^*)^{k+1}\|\boldsymbol{X}^{(0)}-\boldsymbol{\alpha}\|_\infty$$

因 $$\lim_{k\to\infty}(\mu^*)^{k+1}=0，故 \lim_{k\to\infty}\|\boldsymbol{X}^{(k+1)}-\boldsymbol{\alpha}\|_\infty=0$$

证得赛德尔迭代法收敛。

下面推导误差估计公式。由于

$$x_i^{(k+1)} = \sum_{j=1}^{i-1} m_{ij}x_j^{(k+1)} + \sum_{j=i}^{n} m_{ij}x_j^{(k)} + n_i$$

$$x_i^{(k)} = \sum_{j=1}^{i-1} m_{ij}x_j^{(k)} + \sum_{j=i}^{n} m_{ij}x_j^{(k-1)} + n_i$$

两式相减得

$$x_i^{(k+1)} - x_i^{(k)} = \sum_{j=1}^{i-1} m_{ij}(x_j^{(k+1)}-x_j^{(k)}) + \sum_{j=i}^{n} m_{ij}(x_j^{(k)}-x_j^{(k-1)}) \tag{4.80}$$

仿以上相同的推演过程可得

$$\|\boldsymbol{X}^{(k+1)}-\boldsymbol{X}^{(k)}\|_\infty \leqslant \mu^*\|\boldsymbol{X}^{(k)}-\boldsymbol{X}^{(k-1)}\|_\infty \tag{4.81}$$

使用关系式(4.81)，可以得到

$$\begin{aligned}
\|\boldsymbol{X}^{(k+p)}-\boldsymbol{X}^{(k)}\|_\infty &\leqslant \|\boldsymbol{X}^{(k+p)}-\boldsymbol{X}^{(k+p-1)}\|_\infty + \|\boldsymbol{X}^{(k+p-1)}-\boldsymbol{X}^{(k+p-2)}\|_\infty + \cdots + \|\boldsymbol{X}^{(k+1)}-\boldsymbol{X}^{(k)}\|_\infty \\
&\leqslant (\mu^*)^p\|\boldsymbol{X}^{(k)}-\boldsymbol{X}^{(k-1)}\|_\infty + (\mu^*)^{p-1}\|\boldsymbol{X}^{(k)}-\boldsymbol{X}^{(k-1)}\|_\infty + \cdots + \mu^*\|\boldsymbol{X}^{(k)}-\boldsymbol{X}^{(k-1)}\|_\infty \\
&= [(\mu^*)^p + (\mu^*)^{p-1} + \cdots + \mu^*]\|\boldsymbol{X}^{(k)}-\boldsymbol{X}^{(k-1)}\|_\infty \\
&\leqslant \frac{\mu^*}{1-\mu^*}\|\boldsymbol{X}^{(k)}-\boldsymbol{X}^{(k-1)}\|_\infty
\end{aligned} \tag{4.82}$$

因迭代过程收敛，当 $p\to\infty$ 时，就有

$$\lim_{p\to\infty}\boldsymbol{X}^{(k+p)} = \boldsymbol{\alpha}$$

对式(4.82)取极限得

$$\|\boldsymbol{X}^{(k)}-\boldsymbol{\alpha}\|_\infty \leqslant \frac{\mu^*}{1-\mu^*}\|\boldsymbol{X}^{(k)}-\boldsymbol{X}^{(k-1)}\|_\infty \tag{4.83}$$

对上式反复使用关系式(4.81)得

$$\begin{aligned}
\|\boldsymbol{X}^{(k)}-\boldsymbol{\alpha}\|_\infty &\leqslant \frac{\mu^*}{1-\mu^*}\|\boldsymbol{X}^{(k)}-\boldsymbol{X}^{(k-1)}\|_\infty \\
&\leqslant \frac{(\mu^*)^2}{1-\mu}\|\boldsymbol{X}^{(k-1)}-\boldsymbol{X}^{(k-2)}\|_\infty \\
&\leqslant \cdots
\end{aligned}$$

$$\leqslant \frac{(\mu^*)^k}{1-\mu^*}\|X^{(1)}-X^{(0)}\|_\infty$$ （证毕）

### 3.2.3　充分条件 2

若 $\nu=\|M\|_1<1$，则对任意初值，赛德尔迭代法收敛，且

$$\|X^{(k)}-\alpha\|_1 \leqslant \frac{\rho^k}{(1-s)(1-\rho)}\|X^{(1)}-X^{(0)}\|_1 \tag{4.84}$$

其中　$t_j=\sum_{i=1}^{j}|m_{ij}|,\quad s_j=\sum_{i=j+1}^{n}|m_{ij}|,\quad s=\max_j s_j,\quad \rho=\max_j\frac{t_j}{1-s_j}$

上式中的 $t_j$，$s_j$ 及 $s$ 如图 4.2 所示。

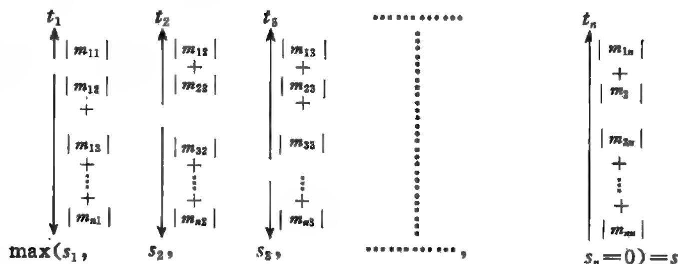


图 4.2

**证**　对公式(4.73)求和得

$$\sum_{i=1}^{n}|x_i^{(k+1)}-\alpha_i|\leqslant\sum_{i=1}^{n}\sum_{j=1}^{i-1}|m_{ij}||x_j^{(k+1)}-\alpha_j|+\sum_{i=1}^{n}\sum_{j=i}^{n}|m_{ij}||x_j^{(k)}-\alpha_j| \tag{4.85}$$

在上式右边更换求和次序得

$$\sum_{i=1}^{n}|x_i^{(k+1)}-\alpha_i|\leqslant\sum_{j=1}^{n}|x_j^{(k+1)}-\alpha_j|\sum_{i=j+1}^{n}|m_{ij}|+\sum_{j=1}^{n}|x_j^{(k)}-\alpha_j|\sum_{i=1}^{j}|m_{ij}| \tag{4.86}$$

令　$t_j=\sum_{i=1}^{j}|m_{ij}|,\quad s_j=\sum_{i=j+1}^{n}|m_{ij}|,\quad s_n=|m_{(n+1)n}|=0\qquad (j=1,2,\cdots,n)$

显然
$$\begin{cases}t_j+s_j=\sum_{i=1}^{n}|m_{ij}|\leqslant\nu<1\\0\leqslant s_j<1\end{cases} \tag{4.87}$$

则式(4.86)可表为

$$\sum_{i=1}^{n}|x_i^{(k+1)}-\alpha_i|\leqslant\sum_{j=1}^{n}s_j|x_j^{(k+1)}-\alpha_j|+\sum_{j=1}^{n}t_j|x_j^{(k)}-\alpha_j|$$

或
$$\sum_{j=1}^{n}(1-s_j)|x_j^{(k+1)}-\alpha_j|\leqslant\sum_{j=1}^{n}t_j|x_j^{(k)}-\alpha_j| \tag{4.88}$$

因
$$t_j\leqslant\nu-s_j\leqslant\nu-\nu s_j=\nu(1-s_j) \tag{4.89}$$

则可把式(4.88)化为以下递推式

$$\sum_{j=1}^{n}(1-s_j)|x_j^{(k+1)}-\alpha_j|\leqslant\nu\sum_{j=1}^{n}(1-s_j)|x_j^{(k)}-\alpha_j| \tag{4.90}$$

反复使用以上关系式得

$$\sum_{j=1}^{n}(1-s_j)|x_j^{(k+1)}-\alpha_j|\leqslant\nu\sum_{j=1}^{n}(1-s_j)|x_j^{(k)}-\alpha_j|$$

$$\leqslant \nu^2 \sum_{j=1}^{n}(1-s_j)\left|x_j^{(k-1)}-\alpha_j\right|$$

$$\cdots$$

$$\leqslant \nu^{k+1} \sum_{j=1}^{n}(1-s_j)\left|x_j^{(0)}-\alpha_j\right| \tag{4.91}$$

因 $\nu<1, 1-s_j\neq 0$，故 $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n}(1-s_j)\left|x_j^{(k+1)}-\alpha_j\right|=0$

证得 $\lim\limits_{k\to\infty} x_j^{(k+1)}=\alpha_j\,(j=1,2,\cdots,n)$，即赛德尔迭代法收敛。

同法可以得到

$$\begin{aligned}\sum_{j=1}^{n}(1-s_j)\left|x_j^{(k+1)}-x_j^{(k)}\right| &\leqslant \sum_{j=1}^{n}t_j\left|x_j^{(k)}-x_j^{(k-1)}\right| \\ &=\sum_{j=1}^{n}\frac{t_j}{1-s_j}(1-s_j)\left|x_j^{(k)}-x_j^{(k-1)}\right| \\ &\leqslant \max_j\frac{t_j}{1-s_j}\sum_{j=1}^{n}(1-s_j)\left|x_j^{(k)}-x_j^{(k-1)}\right| \\ &=\rho\sum_{j=1}^{n}(1-s_j)\left|x_j^{(k)}-x_j^{(k-1)}\right|\end{aligned} \tag{4.92}$$

其中

$$\rho=\max_j\frac{t_j}{1-s_j}$$

因 $t_j\leqslant \nu-s_j\leqslant \nu-\nu s_j=\nu(1-s_j)$，即得

$$\frac{t_j}{1-s_j}\leqslant \nu<1 \qquad (j=1,2,\cdots,n)$$

推知 $\rho<1$。记 $\sigma_{k+1}=\sum\limits_{j=1}^{n}(1-s_j)\left|x_j^{(k+1)}-x_j^{(k)}\right|$，则式(4.92)可化为

$$\sigma_{k+1}\leqslant \rho\,\sigma_k \tag{4.93}$$

反复使用上式，可得

$$\sigma_{k+p}\leqslant \rho\,\sigma_{k+p-1}\leqslant \rho^2\sigma_{k+p-2}\leqslant\cdots\leqslant \rho^p\sigma_k \qquad (p=1,2,\cdots) \tag{4.94}$$

继续利用上式可推出以下不等式

$$\begin{aligned}\sum_{j=1}^{n}(1-s_j)\left|x_j^{k+p}-x_j^{(k)}\right| &\leqslant \sigma_{k+p}+\sigma_{k+p-1}+\cdots+\sigma_{k+1} \\ &\leqslant \rho^p\sigma_k+\rho^{p-1}\sigma_k+\cdots+\rho\,\sigma_k \\ &=(\rho^p+\rho^{p-1}+\cdots+\rho)\sigma_k \\ &\leqslant \frac{\rho}{1-\rho}\sigma_k\end{aligned} \tag{4.95}$$

因迭代过程收敛，所以

$$\lim_{p\to\infty}x_j^{(k+p)}=\alpha_j$$

对式(4.95)取极限得

$$\lim_{p\to\infty}\sum_{j=1}^{n}(1-s_j)\left|x_j^{(k+p)}-x_j^{(k)}\right|=\sum_{j=1}^{n}(1-s_j)\left|x_j^{(k)}-\alpha_j\right|\leqslant\frac{\rho}{1-\rho}\sigma_k \tag{4.96}$$

记

$$s=\max_j s_j$$

因 $0\leqslant s_j<1(j=1,2,\cdots,n)$，$s_j\leqslant s$，则有以下不等式成立

$$1-s\leqslant 1-s_j$$

$$(1-s)\sum_{j=1}^{n}|x_j^{(k)}-\alpha_j|\leqslant\sum_{j=1}^{n}(1-s_j)|x_j^{(k)}-\alpha_j|\leqslant\frac{\rho}{1-\rho}\sigma_k \tag{4.97}$$

$$\begin{aligned}(1-s)\sum_{j=1}^{n}|x_j^{(k)}-\alpha_j| &\leqslant\frac{\rho}{1-\rho}\sigma_k\\ &\leqslant\frac{\rho^2}{1-\rho}\sigma_{k-1}\\ &\cdots\\ &\leqslant\frac{\rho^k}{1-\rho}\sigma_1=\frac{\rho^k}{1-\rho}\sum(1-s_j)|x_j^{(1)}-x_j^{(0)}|\\ &\leqslant\frac{\rho^k}{1-\rho}\sum_{j=1}^{n}|x_j^{(1)}-x_j^{(0)}|\end{aligned}$$

因得
$$\sum_{j=1}^{n}|x_j^{(k)}-\alpha_j|\leqslant\frac{\rho^k}{(1-s)(1-\rho)}\sum_{j=1}^{n}|x_j^{(1)}-x_j^{(0)}|$$

或
$$\|\boldsymbol{X}^{(k)}-\boldsymbol{\alpha}\|_1\leqslant\frac{\rho^k}{(1-s)(1-\rho)}\|\boldsymbol{X}^{(1)}-\boldsymbol{X}^{(0)}\|_1$$

（证毕）

### 3.2.4　充分条件 3

若 $p=\|\boldsymbol{M}\|_F<1$，则对任意初值，赛德尔迭代法收敛，且

$$\|\boldsymbol{X}^{(k)}-\boldsymbol{\alpha}\|_2^2\leqslant\frac{\rho^k}{(1-s)(1-\rho)}\|\boldsymbol{X}^{(1)}-\boldsymbol{X}^{(0)}\|_2^2 \tag{4.98}$$

其中
$$\theta_i=\sum_{j=1}^{n}m_{ij}^2\qquad(i=1,2,\cdots,n)$$

$$t_j=\sum_{i=1}^{j}\theta_i,\quad s_j=\sum_{i=j+1}^{n}\theta_i,\quad s=\max_j s_j,\quad \rho=\max_j\frac{t_j}{1-s_j}$$

上式中的 $\theta_i,t_j,s_j$ 及 $s$ 如图 4.3 所示。

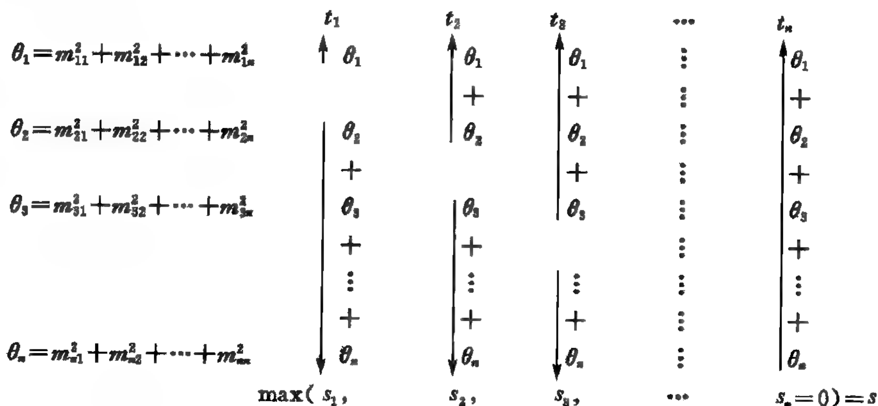



图 4.3

**证**　对式(4.73)两边平方得

$$|x_i^{(k+1)}-\alpha_i|^2 \leqslant \left[\sum_{j=1}^{i-1}|m_{ij}|\,|x_j^{(k+1)}-\alpha_j|+\sum_{j=i}^{n}|m_{ij}|\,|x_j^{(k)}-\alpha_j|\right]^2 \tag{4.99}$$

应用柯西不等式

$$\left|\sum_{s=1}^{n}a_s b_s\right|^2 \leqslant \left(\sum_{s=1}^{n}|a_s|^2\right)\cdot\left(\sum_{s=1}^{n}|b_s|^2\right) \tag{4.100}$$

得
$$\begin{aligned}|x_i^{(k+1)}-\alpha_i|^2 &\leqslant \left(\sum_{j=1}^{n}|m_{ij}|^2\right)\cdot\left(\sum_{j=1}^{i-1}|x_j^{(k+1)}-\alpha_j|^2+\sum_{j=i}^{n}|x_j^{(k)}-\alpha_j|^2\right)\\ &=\theta_i\left(\sum_{j=1}^{i-1}|x_j^{(k+1)}-\alpha_j|^2+\sum_{j=i}^{n}|x_j^{(k)}-\alpha_j|^2\right)\qquad (i=1,2,\cdots,n)\end{aligned} \tag{4.101}$$

其中
$$\theta_i=\sum_{j=1}^{n}m_{ij}^2$$

对式(4.101)从 $i=1,2,\cdots,n$ 求和得

$$\sum_{i=1}^{n}|x_i^{(k+1)}-\alpha_i|^2 \leqslant \sum_{i=1}^{n}\sum_{j=1}^{i-1}\theta_i|x_j^{(k+1)}-\alpha_j|^2+\sum_{i=1}^{n}\sum_{j=i}^{n}\theta_i|x_j^{(k)}-\alpha_j|^2 \tag{4.102}$$

在上式中,将左边的 $i$ 换为 $j$ 而在右边更换求和次序得

$$\begin{aligned}\sum_{j=1}^{n}|x_j^{(k+1)}-\alpha_j|^2 &\leqslant \sum_{j=1}^{n}|x_j^{(k+1)}-\alpha_j|^2\sum_{i=j+1}^{n}\theta_i+\sum_{j=1}^{n}|x_j^{(k)}-\alpha_j|^2\sum_{i=1}^{j}\theta_i\\ &=\sum_{j=1}^{n}s_j|x_j^{(k+1)}-\alpha_j|^2+\sum_{j=1}^{n}t_j|x_j^{(k)}-\alpha_j|^2\end{aligned} \tag{4.103}$$

其中
$$t_j=\sum_{i=1}^{j}\theta_i,\quad s_j=\sum_{i=j+1}^{n}\theta_i$$

显然
$$t_j+s_j=\sum_{i=1}^{n}\theta_i=\sum_{i,j=1}^{n}m_{ij}^2=\|M\|_F^2=p^2<1$$

$$t_j=p^2-s_j\leqslant p^2-p^2s_j=p^2(1-s_j) \tag{4.104}$$

则式(4.103)可化为以下递推关系

$$\sum_{j=1}^{n}(1-s_j)|x_j^{(k+1)}-\alpha_j|^2 \leqslant \sum_{j=1}^{n}t_j|x_j^{(k)}-\alpha_j|^2 \leqslant p^2\sum_{j=1}^{n}(1-s_j)|x_j^{(k)}-\alpha_j|^2 \tag{4.105}$$

反复利用式(4.105)得

$$\begin{aligned}\sum_{j=1}^{n}(1-s_j)|x_j^{(k+1)}-\alpha_j|^2 &\leqslant p^2\sum_{j=1}^{n}(1-s_j)|x_j^{(k)}-\alpha_j|^2\\ &\leqslant (p^2)^2\sum_{j=1}^{n}(1-s_j)|x_j^{(k-1)}-\alpha_j|^2\\ &\cdots\\ &\leqslant (p^2)^{k+1}\sum_{j=1}^{n}(1-s_j)|x_j^{(0)}-\alpha_j|^2\end{aligned} \tag{4.106}$$

因 $p^2<1,1-s_j\neq0$,对上式取极限得

$$\lim_{k\to\infty}\sum_{j=1}^{n}(1-s_j)|x_j^{(k+1)}-\alpha_j|^2=0$$

证得
$$\lim_{k\to\infty} x_j^{(k+1)} = \alpha_j$$
所以赛德尔迭代法收敛。

对于误差估计公式(4.98)可按推证误差估计公式(4.84)类似过程证得。

对于高斯-赛德尔迭代法的收敛性有以下定理描述。

**定理 4.10** 若 $\boldsymbol{A}$ 为不可约、对角占优矩阵，则高斯-赛德尔迭代法必定收敛。

**证** 要证明高斯-赛德尔迭代法收敛，根据定理 4.6，只要证明 $\rho(\boldsymbol{G})<1$ 即可，$\boldsymbol{G}$ 是高斯-赛德尔迭代法的迭代矩阵。

由式(4.65)知，高斯-赛德尔迭代法的迭代公式为
$$\boldsymbol{X}^{(k+1)} = -\boldsymbol{D}^{-1}\boldsymbol{L}\boldsymbol{X}^{(k+1)} - \boldsymbol{D}^{-1}\boldsymbol{U}\boldsymbol{X}^{(k)} + \boldsymbol{D}^{-1}\boldsymbol{B}$$
可将它化为等价的简单迭代法形式：
$$(\boldsymbol{I}+\boldsymbol{D}^{-1}\boldsymbol{L})\boldsymbol{X}^{(k+1)} = -\boldsymbol{D}^{-1}\boldsymbol{U}\boldsymbol{X}^{(k)} + \boldsymbol{D}^{-1}\boldsymbol{B}$$
$$\begin{aligned}\boldsymbol{X}^{(k+1)} &= -(\boldsymbol{I}+\boldsymbol{D}^{-1}\boldsymbol{L})^{-1}\boldsymbol{D}^{-1}\boldsymbol{U}\boldsymbol{X}^{(k)} + (\boldsymbol{I}+\boldsymbol{D}^{-1}\boldsymbol{L})\boldsymbol{D}^{-1}\boldsymbol{B}\\ &= -[\boldsymbol{D}(\boldsymbol{I}+\boldsymbol{D}^{-1}\boldsymbol{L})]^{-1}\boldsymbol{U}\boldsymbol{X}^{(k)} + (\boldsymbol{I}+\boldsymbol{D}^{-1}\boldsymbol{L})^{-1}\boldsymbol{D}^{-1}\boldsymbol{B}\\ &= -(\boldsymbol{D}+\boldsymbol{L})^{-1}\boldsymbol{U}\boldsymbol{X}^{(k)} + (\boldsymbol{I}+\boldsymbol{D}^{-1}\boldsymbol{L})^{-1}\boldsymbol{D}^{-1}\boldsymbol{B}\end{aligned} \tag{4.107}$$
因此高斯-赛德尔迭代法的迭代矩阵为
$$\boldsymbol{G} = -(\boldsymbol{D}+\boldsymbol{L})^{-1}\boldsymbol{U} \tag{4.108}$$
下面用反证法推证定理。假设 $\boldsymbol{G}$ 的特征值有 $|\lambda|\geqslant 1$，则它必满足以下特征方程
$$|\lambda\boldsymbol{I}-\boldsymbol{G}| = 0 \tag{4.109}$$
将式(4.108)代入上式得
$$\begin{aligned}|\lambda\boldsymbol{I} + (\boldsymbol{D}+\boldsymbol{L})^{-1}\boldsymbol{U}| &= 0\\ |(\boldsymbol{D}+\boldsymbol{L})^{-1}[\lambda(\boldsymbol{D}+\boldsymbol{L})+\boldsymbol{U}]| &= 0\\ |(\boldsymbol{D}+\boldsymbol{L})^{-1}|\,|\lambda(\boldsymbol{D}+\boldsymbol{L})+\boldsymbol{U}| &= 0\end{aligned} \tag{4.110}$$
由于 $\boldsymbol{A}$ 不可约、且具有对角占优，所以 $a_{ii}\neq 0(i=1,2,\cdots,n)$，因此有 $|(\boldsymbol{D}+\boldsymbol{L})^{-1}|\neq 0$，由式(4.110)推知
$$|\tilde{\boldsymbol{G}}| = |\lambda(\boldsymbol{D}+\boldsymbol{L})+\boldsymbol{U}| = 0 \tag{4.111}$$
其中矩阵 $\tilde{\boldsymbol{G}}$ 为
$$\tilde{\boldsymbol{G}} = \lambda(\boldsymbol{D}+\boldsymbol{L})+\boldsymbol{U} = \begin{bmatrix}\lambda a_{11} & a_{12} & a_{13} & \cdots & a_{1n}\\ \lambda a_{21} & \lambda a_{22} & a_{23} & \cdots & a_{2n}\\ \cdots & & \ddots & & \cdots\\ & & & & a_{(n-1)n}\\ \lambda a_{n1} & & \cdots & & \lambda a_{nn}\end{bmatrix} \tag{4.112}$$
由于矩阵 $\tilde{\boldsymbol{G}}$ 中的零元素的位置与矩阵 $\boldsymbol{A}$ 中零元素的位置全同，由 $\boldsymbol{A}$ 的不可约性即可推得 $\tilde{\boldsymbol{G}}$ 的不可约性。再由 $\boldsymbol{A}$ 具有对角占优得
$$|a_{ii}| \geqslant \sum_{j=1}^{i-1}|a_{ij}| + \sum_{j=i+1}^{n}|a_{ij}|$$
两边同乘 $|\lambda|$ 得
$$|\lambda a_{ii}| \geqslant \sum_{j=1}^{i-1}|\lambda a_{ij}| + \sum_{j=i+1}^{n}|\lambda a_{ij}|$$

$$\geqslant \sum_{j=1}^{i-1} |\lambda a_{ij}| + \sum_{j=i+1}^{n} |a_{ij}| \qquad (因 |\lambda| \geqslant 1)$$

可见 $\tilde{G}$ 也是不可约、对角占优矩阵。据引理知 $|\tilde{G}| \neq 0$，与式(4.111)矛盾，故所有 $|\lambda|<1$，即 $\rho(G)<1$。定理得证。

**定理 4.11** 若 $A$ 对称正定，则高斯-赛德尔迭代法收敛。

这个定理是下面将会证明的定理 4.13 的一部分。

应该指出，每一种迭代法都有一定的适用范围。因此，有些线性方程组使用简单迭代法收敛，赛德尔迭代法不收敛。反之，有些线性方程组，赛德尔迭代法收敛，而简单迭代法不收敛。即使在都收敛的情况下，其收敛的快慢也可能不一样。

**例 4.3** 讨论求解线性方程组 $AX=B$ 迭代法的收敛性，其中

$$A = \begin{bmatrix} 1 & -2 & 2 \\ -1 & 1 & -1 \\ -2 & -2 & 1 \end{bmatrix}$$

**解** 当采用雅可比迭代法时，其迭代矩阵的特征方程为

$$\begin{vmatrix} \lambda & -2 & 2 \\ -1 & \lambda & -1 \\ -2 & -2 & \lambda \end{vmatrix} = 0$$

解得 $\lambda_1=\lambda_2=\lambda_3=0$，雅可比迭代法收敛。

当采用高斯-赛德尔迭代法时，其迭代矩阵的特征方程为

$$\begin{vmatrix} \lambda & -2 & 2 \\ -\lambda & \lambda & -1 \\ -2\lambda & -2\lambda & \lambda \end{vmatrix} = 0$$

解得 $\lambda_1=0, \lambda_2=-2+\sqrt{8}, \lambda_3=-2-\sqrt{8}$。其谱半径为 $2+\sqrt{8}>1$，高斯-赛德尔迭代法不收敛。

**例 4.4** 讨论求解线性方程组 $AX=B$ 迭代法的收敛性，其中

$$A = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & -2 \end{bmatrix}$$

**解** 当采用雅可比迭代法时，其迭代矩阵的特征方程为

$$\begin{vmatrix} 2\lambda & -1 & 1 \\ 1 & \lambda & 1 \\ 1 & 1 & -2\lambda \end{vmatrix} = 0$$

解得 $\lambda_1=0, \lambda_2=\frac{\sqrt{5}}{2}i, \lambda_3=-\frac{\sqrt{5}}{2}i$。其谱半径为 $\frac{\sqrt{5}}{2}>1$，雅可比迭代法不收敛。

当采用高斯-赛德尔迭代法时，其迭代矩阵的特征方程是

$$\begin{vmatrix} 2\lambda & -1 & 1 \\ \lambda & \lambda & 1 \\ \lambda & \lambda & -2\lambda \end{vmatrix} = 0$$

解得 $\lambda_1=0, \lambda_2=\lambda_3=-0.5$。其谱半径为 $0.5<1$，高斯-赛德尔迭代法收敛。

# §4 松弛迭代法

记线性方程组的残差为

$$R_i=b_i-(a_{i1}x_1+a_{i2}x_2+\cdots+a_{in}x_n)\ (i=1,2,\cdots,n) \tag{4.113}$$

为便于进行松弛计算，我们把上述残差 $R_i$ 改写成

$$\begin{aligned} r_i &=\frac{R_i}{a_{ii}}=-x_i-\frac{1}{a_{ii}}(a_{i1}x_1+\cdots+a_{ii-1}x_{i-1}+a_{ii+1}x_{i+1}+\cdots+a_{in}x_n)+\frac{b_i}{a_{ii}} \\ &=-x_i+(b_{i1}x_1+\cdots+b_{ii-1}x_{i-1}+b_{ii+1}x_{i+1}+\cdots+b_{in}x_n)+f_i \\ &=-x_i+\sum_{\substack{j=1\\j\neq i}}^{n}b_{ij}x_j+f_i \quad (i=1,2,\cdots,n) \end{aligned} \tag{4.114}$$

式中，$b_{ij}=-\dfrac{a_{ij}}{a_{ii}}\quad(i\neq j),f_i=\dfrac{b_i}{a_{ii}}$。

根据对方程松弛顺序的不同控制策略，我们叙述以下几种松弛迭代法。

## 4.1 按 $\max\limits_l|r_l^{(k)}|$ 实施松弛的松弛法

设当 $\boldsymbol{X}=\boldsymbol{X}^{(k)}$ 时由(4.114)式所确定的残差为 $(r_1^{(k)},r_2^{(k)},\cdots,r_n^{(k)})'$，在本法中，依据 $r_i^{(k)}=\max\limits_l|r_l^{(k)}|(l=1,2,\cdots,n)$ 所确定的 $i$，将第 $i$ 个方程中的 $x_i^{(k)}$ 修改为 $\hat{x}_i$，使第 $i$ 个方程的残差 $r_i^{(k+1)}\equiv0$，由式(4.114)推得

$$0\equiv r_i^{(k+1)}=-\hat{x}_i+\sum_{\substack{j=1\\j\neq i}}^{n}b_{ij}x_j^{(k)}+f_i \tag{4.115}$$

$$\begin{aligned} \hat{x}_i &=\sum_{\substack{j=1\\j\neq i}}^{n}b_{ij}x_j^{(k)}+f_i=-\frac{1}{a_{ii}}(a_{i1}x_1^{(k)}+\cdots+a_{ii-1}x_{i-1}^{(k)}+a_{ii+1}x_{i+1}^{(k)}+\cdots a_{in}x_n^{(k)})+\frac{b_i}{a_{ii}} \\ &=x_i^{(k)}+\frac{1}{a_{ii}}[b_i-(a_{i1}x_1^{(k)}+\cdots+a_{ii-1}x_{i-1}^{(k)}+a_{ii}x_i^{(k)}+a_{ii+1}x_{i+1}^{(k)}+\cdots+a_{in}x_n^{(k)})] \\ &=x_i^{(k)}+\frac{1}{a_{ii}}R_i^{(k)}=x_i^{(k)}+r_i^{(k)} \end{aligned} \tag{4.116}$$

取 $\boldsymbol{X}^{(k+1)}=(x_1^{(k)},\cdots,x_{i-1}^{(k)},\hat{x}_i,x_{i+1}^{(k)},\cdots,x_n^{(k)})'$，继续求取 $r^{(k+1)}$，仿上推作，直至近似值的全部修改量均小于给定精度要求为止。

一般，可建立具有松弛因子 $\widetilde{w}$ 的修改公式如下

$$x_i^{(k+1)}=x_i^{(k)}+\widetilde{w}r_i^{(k)}=x_i^{(k)}+\widetilde{w}(\hat{x}_i-x_i^{(k)}) \tag{4.117}$$

当 $\widetilde{w}=1$ 时，式(4.117)转化为式(4.116)。

## 4.2 简单迭代方式下的逐次松弛法

在本法中，依据方程排列的顺序 $i=1,2,\cdots,n$ 修改 $x_i^{(k)}$ 为 $\hat{x}_i$ 使 $r_i^{(k+1)}\equiv0$，由式(4.116)可得 $\hat{x}_i=x_i^{(k)}+r_i^{(k)}(i=1,2,\cdots,n)$，令 $x_i^{(k+1)}=\hat{x}_i$，则得简单迭代方式下的逐次松弛迭代公式

$$x_i^{(k+1)}=x_i^{(k)}+r_i^{(k)}\quad(i=1,2,\cdots,n) \tag{4.118}$$

由式(4.118)可见，它完全等同于雅可比迭代法式(4.32)。

一般，可建立具有松弛因子 $\widetilde{w}$ 的修改公式如下

$$x_i^{(k+1)}=x_i^{(k)}+\widetilde{w}r_i^{(k)}=x_i^{(k)}+\frac{\widetilde{w}}{a_{ii}}R_i^{(k)}\quad(i=1,2,\cdots,n)\tag{4.119}$$

当 $\widetilde{w}=1$ 时,(4.119)式转化为(4.118)式。

**例 4.5** 使用公式(4.119)建立下述线性方程组

$$\begin{cases}4x_1+3x_2=24\\3x_1+4x_2-x_3=30\\-x_2+4x_3=-24\end{cases}\tag{4.120}$$

的松弛迭代公式。

**解** 按式(4.119)可得如下松弛迭代公式

$$\begin{cases}x_1^{(k+1)}=x_1^{(k)}+\dfrac{\widetilde{w}}{4}(24-4x_1^{(k)}-3x_2^{(k)})\\x_2^{(k+1)}=x_2^{(k)}+\dfrac{\widetilde{w}}{4}(30-3x_1^{(k)}-4x_2^{(k)}+x_3^{(k)})\\x_3^{(k+1)}=x_3^{(k)}+\dfrac{\widetilde{w}}{4}(-24+x_2^{(k)}-4x_3^{(k)})\end{cases}\tag{4.121}$$

式(4.119)尚可改写为

$$x_i^{(k+1)}=x_i^{(k)}+\widetilde{w}(\hat{x}_i-x_i^{(k)})=(1-\widetilde{w})x_i^{(k)}+\widetilde{w}\hat{x}_i\quad(i=1,2,\cdots,n)\tag{4.122}$$

其中

$$\hat{x}_i=-\frac{1}{a_{ii}}(a_{i1}x_1^{(k)}+\cdots+a_{ii-1}x_{i-1}^{(k)}+a_{ii+1}x_{i+1}^{(k)}+\cdots+a_{in}x_n^{(k)}-b_i)$$

为使用雅可比迭代法所得的近似值。由式(4.122)可见,带有松弛因子 $\widetilde{w}$ 按简单迭代方式的逐次松弛法就是雅可比迭代法中新、旧两个迭代值 $\hat{x}_i$、$x_i^{(k)}$ 依 $\widetilde{w}$、$(1-\widetilde{w})$ 加权的组合公式。

## 4.3 赛德尔迭代方式下的逐次松弛法

类似地可导得赛德尔迭代方式下的逐次松弛迭代公式

$$x_i^{(k+1)}=-\frac{1}{a_{ii}}(a_{i1}x_1^{(k+1)}+\cdots+a_{ii-1}x_{i-1}^{(k+1)}+a_{ii+1}x_{i+1}^{(k)}+\cdots+a_{in}x_n^{(k)}-b_i)\quad(i=1,2,\cdots,n)\tag{4.123}$$

由式(4.123)可见,它完全等同于高斯—赛德尔法式(4.64)。

一般,可建立具有松弛因子 $\widetilde{w}$ 的修改公式如下

$$\begin{aligned}x_i^{(k+1)}&=-\frac{\widetilde{w}}{a_{ii}}(a_{i1}x_1^{(k+1)}+\cdots+a_{ii-1}x_{i-1}^{(k+1)}+a_{ii+1}x_{i+1}^{(k)}+\cdots+a_{in}x_n^{(k)}-b_i)\\&=x_i^{(k)}+\frac{\widetilde{w}}{a_{ii}}[b_i-(a_{i1}x_1^{(k+1)}+\cdots+a_{ii-1}x_{i-1}^{(k+1)}+a_{ii}x_i^{(k)}+a_{ii+1}x_{i+1}^{(k)}+\cdots+a_{in}x_n^{(k)})]\\&=x_n^{(k)}+\widetilde{w}(\hat{x}_i-x_i^{(k)})=(1-\widetilde{w})x_i^{(k)}+\widetilde{w}\hat{x}_i\end{aligned}\tag{4.124}$$

其中

$$\hat{x}_i=-\frac{1}{a_{ii}}(a_{i1}x_1^{(k+1)}+\cdots+a_{ii-1}x_{i-1}^{(k+1)}+a_{ii+1}x_{i+1}^{(k)}+\cdots+a_{in}x_n^{(k)}-b_i)$$

为使用高斯—赛德尔法所得的近似值,由式(4.124)可见,带有松弛因子 $\widetilde{w}$ 按赛德尔迭代方式的逐次松弛法就是高斯—赛德尔迭代中新、旧两个迭代值 $\hat{x}_i$、$x_i^{(k)}$ 依 $\widetilde{w}$、$(1-\widetilde{w})$ 加权的组合公

式。当 $\widetilde{w}=1$ 时，式(4.124)转化为式(4.123)。

**例 4.6**　用松弛法式(4.124)解下列方程组

$$\begin{cases} 4x_1+3x_2=24 \\ 3x_1+4x_2-x_3=30 \\ -x_2+4x_3=-24 \end{cases}$$

**解**　按式(4.124)建立松弛迭代公式

$$\begin{cases} x_1^{(k+1)}=x_1^{(k)}+\dfrac{\widetilde{w}}{4}(24-4x_1^{(k)}-3x_2^{(k)}) \\ x_2^{(k+1)}=x_2^{(k)}+\dfrac{\widetilde{w}}{4}(30-3x_1^{(k+1)}-4x_2^{(k)}+x_3^{(k)}) \\ x_3^{(k+1)}=x_3^{(k)}+\dfrac{\widetilde{w}}{4}(-24+x_2^{(k+1)}-4x_3^{(k)}) \end{cases} \tag{4.125}$$

当 $\widetilde{w}=1$ 时，即得高斯-赛德尔迭代公式

$$\begin{cases} x_1^{(k+1)}=-0.75x_2^{(k)}+6 \\ x_2^{(k+1)}=-0.75x_1^{(k+1)}+0.25x_3^{(k)}+7.5 \\ x_3^{(k+1)}=0.25x_2^{(k+1)}-6 \end{cases} \tag{4.126}$$

当 $\widetilde{w}=1.25$ 时，得以下松弛迭代公式

$$\begin{cases} x_1^{(k+1)}=-0.25x_1^{(k)}-0.9375x_2^{(k)}+7.5 \\ x_2^{(k+1)}=-0.9375x_1^{(k+1)}-0.25x_2^{(k)}+0.3125x_3^{(k)}+9.375 \\ x_3^{(k+1)}=0.3125x_2^{(k+1)}-0.25x_3^{(k)}-7.5 \end{cases} \tag{4.127}$$

上述两组迭代公式均取初值 $x_1^{(0)}=x_2^{(0)}=x_3^{(0)}=1$，若要迭代结果精确到 7 位小数，高斯-赛德尔迭代法需要 34 次迭代计算；而带有 $\widetilde{w}=1.25$ 的松弛迭代法只需 14 次迭代计算。所以只要选好参数 $\omega$ 的值，松弛法的收敛速度是比较快的。松弛法(4.124)有下述收敛定理。

**定理 4.12**　松弛法式(4.124)收敛的必要条件是

$$0<\widetilde{w}<2$$

**证**　由式(4.124)知，它可表为

$$\boldsymbol{X}^{(k+1)}=(1-\widetilde{w})\boldsymbol{X}^{(k)}+\widetilde{w}[\widetilde{\boldsymbol{L}}\boldsymbol{X}^{(k+1)}+\widetilde{\boldsymbol{U}}\boldsymbol{X}^{(k)}+\boldsymbol{F}] \tag{4.128}$$

其中

$$\widetilde{\boldsymbol{L}}=\begin{bmatrix} 0 & & & \\ -\dfrac{a_{21}}{a_{22}} & 0 & & \text{\Large 0} \\ -\dfrac{a_{31}}{a_{33}} & -\dfrac{a_{32}}{a_{33}} & 0 & \\ \vdots & & \ddots & \ddots \\ -\dfrac{a_{n1}}{a_{nn}} & \cdots & -\dfrac{a_{n(n-1)}}{a_{nn}} & 0 \end{bmatrix},\quad \widetilde{\boldsymbol{U}}=\begin{bmatrix} 0 & -\dfrac{a_{12}}{a_{11}} & -\dfrac{a_{13}}{a_{11}} & \cdots & -\dfrac{a_{1n}}{a_{11}} \\ & 0 & \dfrac{a_{23}}{a_{22}} & \cdots & -\dfrac{a_{2n}}{a_{22}} \\ & & \ddots & \ddots & \vdots \\ & \text{\Large 0} & & & -\dfrac{a_{(n-1)n}}{a_{(n-1)(n-1)}} \\ & & & & 0 \end{bmatrix}$$

$$\boldsymbol{X}^{(k+1)}=\begin{bmatrix}x_1^{(k+1)}\\x_2^{(k+1)}\\\cdots\\x_n^{(k+1)}\end{bmatrix},\quad \boldsymbol{X}^{(k)}=\begin{bmatrix}x_1^{(k)}\\x_2^{(k)}\\\cdots\\x_n^{(k)}\end{bmatrix},\quad \boldsymbol{F}=\begin{bmatrix}\dfrac{b_1}{a_{11}}\\\dfrac{b_2}{a_{22}}\\\cdots\\\dfrac{b_n}{a_{nn}}\end{bmatrix}$$

式(4.128)可转化为等价的简单迭代法形式

$$\boldsymbol{X}^{(k+1)}=(\boldsymbol{I}-\tilde{w}\tilde{\boldsymbol{L}})^{-1}[(1-\tilde{w})\boldsymbol{I}+\tilde{w}\tilde{\boldsymbol{U}}]\boldsymbol{X}^{(k)}+\tilde{w}(\boldsymbol{I}-\tilde{\omega}\tilde{\boldsymbol{L}})^{-1}\boldsymbol{F} \tag{4.129}$$

因此得该松弛法的迭代矩阵为

$$\boldsymbol{S}=(\boldsymbol{I}-\tilde{w}\tilde{\boldsymbol{L}})^{-1}[(1-\tilde{w})\boldsymbol{I}+\tilde{w}\tilde{\boldsymbol{U}}] \tag{4.130}$$

$S$ 的特征方程为

$$|\boldsymbol{S}-\lambda\boldsymbol{I}|=\begin{vmatrix}s_{11}-\lambda & s_{12} & \cdots & s_{1n}\\ s_{21} & s_{22}-\lambda & \cdots & s_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ s_{n1} & s_{n2} & \cdots & s_{nn}-\lambda\end{vmatrix}$$

$$=(-1)^n[\lambda^n-\sigma_1\lambda^{n-1}+\sigma_2\lambda^{n-2}+\cdots+(-1)^n\sigma_n]=0 \tag{4.131}$$

令 $\lambda=0$,知

$$\sigma_n=|\boldsymbol{S}| \tag{4.132}$$

根据代数方程根与系数关系知

$$\sigma_n=\lambda_1\lambda_2\cdots\lambda_n$$

其中 $\lambda_i(i=1,2,\cdots,n)$ 为 $\boldsymbol{S}$ 的特征值。因此有

$$\lambda_1\lambda_2\cdots\lambda_n=|\boldsymbol{S}| \tag{4.133}$$

由于

$$\boldsymbol{I}-\tilde{\omega}\tilde{\boldsymbol{L}}=\begin{bmatrix}1 & & 0\\ & \ddots & \\ \tilde{\omega}\dfrac{a_{ij}}{a_{ii}} & & 1\end{bmatrix}$$

所以 $|\boldsymbol{I}-\tilde{w}\tilde{\boldsymbol{L}}|=1$,即 $|(\boldsymbol{I}-\tilde{w}\tilde{\boldsymbol{L}})^{-1}|=1$。另外

$$(1-\tilde{\omega})\boldsymbol{I}+\tilde{\omega}\tilde{\boldsymbol{U}}=\begin{bmatrix}1-\tilde{\omega} & & -\tilde{\omega}\dfrac{a_{ij}}{a_{ii}}\\ & \ddots & \\ 0 & & 1-\tilde{\omega}\end{bmatrix}$$

所以 $|(1-\tilde{w})\boldsymbol{I}+\tilde{w}\tilde{\boldsymbol{U}}|=(1-\tilde{w})^n$。则有

$$\begin{aligned}|\boldsymbol{S}|&=|(\boldsymbol{I}-\tilde{w}\tilde{\boldsymbol{L}})^{-1}[(1-\tilde{w})\boldsymbol{I}+\tilde{w}\tilde{\boldsymbol{U}}]|\\&=|(\boldsymbol{I}-\tilde{w}\tilde{\boldsymbol{L}})^{-1}|\,|(1-\tilde{w})\boldsymbol{I}+\tilde{w}\tilde{\boldsymbol{U}}|=(1-\tilde{w})^n\end{aligned} \tag{4.134}$$

代入式(4.133)并取绝对值得

$$|(1-\tilde{w})^n| = |\lambda_1\lambda_2\cdots\lambda_n| \tag{4.135}$$

因该松弛法收敛，必有谱半径 $\rho(\boldsymbol{S})<1$，即 $|\lambda_i|<1(i=1,2,\cdots,n)$。所以

$$|(1-\tilde{w})^n| = |1-\tilde{w}|^n = |\lambda_1\lambda_2\cdots\lambda_n| \leqslant |\rho(\boldsymbol{S})^n| < 1$$

即

$$|1-\tilde{w}|<1 \text{ 或 } 0<\tilde{w}<2$$

定理得证。

上述定理说明，若要松弛法收敛，必须选取松弛因子 $\tilde{w}\in(0,2)$。因这是必要条件，所以当松弛因子满足 $0<\tilde{w}<2$ 时，并不能保证所有的松弛法收敛。

**定理 4.13**　若 $\boldsymbol{A}$ 对称正定，则当 $0<\tilde{w}<2$ 时，松弛法恒收敛。

证：设 $\lambda$ 为 $\boldsymbol{S}$ 的特征值，$\boldsymbol{Y}$ 为对应的特征向量，即

$$\boldsymbol{SY} = \lambda\boldsymbol{Y}$$

其中 $\boldsymbol{Y}=(y_1,y_2,\cdots,y_n)\neq 0$。将 $\boldsymbol{S}$ 的表达式(4.130)代入上式得

$$(\boldsymbol{I}-\tilde{w}\tilde{\boldsymbol{L}})^{-1}[(1-\tilde{w})\boldsymbol{I}+\tilde{w}\tilde{\boldsymbol{U}}]\boldsymbol{Y}=\lambda\boldsymbol{Y} \tag{4.136}$$

或

$$[(1-\tilde{w})\boldsymbol{I}+\tilde{w}\tilde{\boldsymbol{U}}]\boldsymbol{Y} = \lambda(\boldsymbol{I}-\tilde{w}\tilde{\boldsymbol{L}})\boldsymbol{Y} \tag{4.137}$$

两边乘以 $\boldsymbol{D}$ 得

$$[(1-\tilde{w})\boldsymbol{D}+\tilde{w}\boldsymbol{D}\tilde{\boldsymbol{U}}]\boldsymbol{Y} = \lambda(\boldsymbol{D}-\tilde{w}\boldsymbol{D}\tilde{\boldsymbol{L}})\boldsymbol{Y} \tag{4.138}$$

式中

$$\boldsymbol{D}\tilde{\boldsymbol{L}} = -\begin{bmatrix} 0 & & & \\ a_{21} & 0 & & \boldsymbol{0} \\ a_{31} & a_{32} & 0 & \\ \vdots & & \ddots & \ddots \\ a_{n-1} & \cdots & a_{n(n-1)} & 0 \end{bmatrix} = -\boldsymbol{L},\ \boldsymbol{D}\tilde{\boldsymbol{U}} = -\begin{bmatrix} 0 & a_{12} & a_{13} & \cdots & a_{1n} \\ & 0 & a_{23} & \cdots & a_{2n} \\ & & \ddots & \ddots & \vdots \\ & \boldsymbol{0} & & & a_{(n-1)n} \\ & & & & 0 \end{bmatrix} = -\boldsymbol{U}$$

所以

$$[(1-\tilde{w})\boldsymbol{D}-\tilde{w}\boldsymbol{U}]\boldsymbol{Y} = \lambda(\boldsymbol{D}+\tilde{w}\boldsymbol{L})\boldsymbol{Y} \tag{4.139}$$

为了分析 $\lambda$ 值，考虑 $\boldsymbol{Y}$ 右乘上式的向量积：

$$(1-\tilde{w})(\boldsymbol{DY},\boldsymbol{Y})-\tilde{w}(\boldsymbol{UY},\boldsymbol{Y}) = \lambda[(\boldsymbol{DY},\boldsymbol{Y})+\tilde{w}(\boldsymbol{LY},\boldsymbol{Y})]$$

解出

$$\lambda = \frac{(1-\tilde{w})(\boldsymbol{DY},\boldsymbol{Y})-\tilde{w}(\boldsymbol{UY},\boldsymbol{Y})}{(\boldsymbol{DY},\boldsymbol{Y})+\tilde{w}(\boldsymbol{LY},\boldsymbol{Y})} \tag{4.140}$$

记 $d=(\boldsymbol{DY},\boldsymbol{Y})$，$l=(\boldsymbol{LY},\boldsymbol{Y})$，因为 $\boldsymbol{A}$ 正定，故 $\boldsymbol{D}$ 亦正定，所以

$$d = (\boldsymbol{DY},\boldsymbol{Y}) > 0$$

已知 $\boldsymbol{A}$ 为对称正定矩阵，即实对称矩阵，此时 $l$ 为实数且 $\boldsymbol{L}'=\boldsymbol{U}$。计算得

$$(\boldsymbol{UY},\boldsymbol{Y}) = (\boldsymbol{L}'\boldsymbol{Y},\boldsymbol{Y}) = (\boldsymbol{Y},\boldsymbol{LY}) = l$$

$$\lambda = \frac{(1-\tilde{w})d-\tilde{w}l}{d+\tilde{w}l} \tag{4.141}$$

建立以下差值

$$\Delta = [(1-\tilde{w})d-\tilde{w}l]^2-(d+\tilde{w}l)^2 = \tilde{w}d(d+2l)(\tilde{w}-2)$$

因 $\boldsymbol{A}$ 正定且 $\boldsymbol{A}=\boldsymbol{D}+\boldsymbol{L}+\boldsymbol{U}$，所以

$$0<(\boldsymbol{AY},\boldsymbol{Y}) = ((\boldsymbol{D}+\boldsymbol{L}+\boldsymbol{U})\boldsymbol{Y},\boldsymbol{Y}) = d+2l$$

因此，当$0<\tilde{w}<2$时，就有$\Delta<0$成立，即$|\lambda|<1$，或$\rho(\boldsymbol{S})<1$成立，从而推得松弛法(4.124)必收敛。

当$\tilde{w}=1$时，上述定理就是定理4.11。定理4.13是充分条件，具有重要的使用价值。

**定理4.14** 若$\boldsymbol{A}$为不可约、对角占优矩阵，且松弛因子$\tilde{w}$满足$0<\tilde{w}\leqslant 1$，则松弛法(4.124)必定收敛。

证明办法和定理4.10类似，这里不再赘述。当$\tilde{w}=1$时，本定理即为定理4.10。

使用松弛法求解线性方程组，关键是要选好松弛因子的数值。所选出的$\omega$值，不仅要使松弛法收敛，而且应有较高的收敛速度。能使松弛法收敛最快的松弛因子叫做最佳松弛因子。关于如何选取最佳松弛因子的问题，目前除少数几种特殊类型的矩阵具有确定最佳松弛因子的理论公式外，对于一般矩阵的最佳松弛因子选择仍没有有效地解决。即使有理论公式的情况下，公式中的有关参数确定也比较困难。实际计算时，通常的办法是选不同的$\tilde{w}$，且$\tilde{w}\in(0,2)$，然后从同一初值出发，用松弛法进行迭代，迭代次数同为$k$次，再比较它们相应的残差或两次近似值间的误差，选取使残差或误差模最小的松弛因子作为最佳松弛因子的近似值。亦可粗取一个松弛因子$\tilde{w}_0\in(0,2)$，然后根据迭代过程收敛的快慢逐步改进(例如用对分法或优选法在(0,2)内逐步选优)，一旦满意后便固定下来继续迭代，以达到加速收敛的目的。

# 习 题 四

4.1　用简单迭代法及赛德尔迭代法解下列方程组。

$$\begin{cases} 20x_1 + 2x_2 + 3x_3 = 24 \\ x_1 + 8x_2 + x_3 = 12 \\ 2x_1 - 3x_2 + 15x_3 = 30 \end{cases}$$

4.2　把下列方程组化成等价的方程组，使之能应用高斯-赛德尔迭代法进行求解，并写出迭代公式。

$$\begin{bmatrix} 64 & -3 & -1 \\ 1 & 1 & 40 \\ 2 & -90 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 14 \\ 20 \\ -5 \end{bmatrix}$$

4.3　讨论利用简单迭代法和赛德尔迭代法解方程 $\boldsymbol{A}_i\boldsymbol{X}=\boldsymbol{B}(i=1,2)$ 时的收敛性。其中

$$\boldsymbol{A}_1 = \begin{bmatrix} 1 & -2 & 2 \\ -1 & 1 & -1 \\ -2 & -2 & 1 \end{bmatrix},\quad \boldsymbol{A}_2 = \begin{bmatrix} 1 & 0.5 & -0.5 \\ -1 & 1 & -1 \\ 0.5 & 0.5 & 1 \end{bmatrix}$$

4.4　试用松弛法(4.124)(取 $\tilde{w}=0.9$)求解方程组

$$\begin{cases} 5x_1 + 2x_2 + x_3 = -12 \\ -x_1 + 4x_2 + 2x_3 = 20 \\ 2x_1 - 3x_2 + 10x_3 = 3 \end{cases}$$

当满足 $\|\boldsymbol{X}^{(k+1)} - \boldsymbol{X}^{(k)}\|_\infty < 10^{-5}$ 时迭代终止，并证明迭代过程是收敛的。

4.5　设系数矩阵

$$\boldsymbol{A} = \begin{bmatrix} a & 1 & 3 \\ 1 & a & 2 \\ -3 & 2 & a \end{bmatrix}$$

问 $a$ 取什么值时，简单迭代法收敛？$a$ 取什么值时，高斯-赛德尔迭代法收敛？

4.6　设线性方程组 $\boldsymbol{AX}=\boldsymbol{B}$ 的系数矩阵为

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & -2 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{bmatrix}$$

证明雅可比迭代法收敛，高斯-赛德尔迭代法发散。

4.7　设线性方程组 $\boldsymbol{AX}=\boldsymbol{B}$ 的系数矩阵为

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0.4 & 0.4 \\ 0.4 & 1 & 0.8 \\ 0.4 & 0.8 & 1 \end{bmatrix}$$

证明高斯-赛德尔迭代法收敛，雅可比迭代法发散。

4.8　求矩阵 $\boldsymbol{Q}$ 的 $\|\boldsymbol{Q}\|_1$，$\|\boldsymbol{Q}\|_2$，$\|\boldsymbol{Q}\|_\infty$，其中

$$Q=\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

4.9 试证对 $n$ 维向量 $\boldsymbol{X}$ 有

$$\|\boldsymbol{X}\|_\infty \leqslant \|\boldsymbol{X}\|_1 \leqslant n\|\boldsymbol{X}\|_\infty$$

# 第五章　插　值　法

实践中常有这样的问题：由实验得到某一函数 $y=f(x)$ 在一系列点 $x_0,x_1,\cdots,x_n$ 处的值 $y_0,y_1,\cdots,y_n$，它们的函数解析式是未知的；或者 $y=f(x)$ 虽有明确的解析式，但计算复杂，不便于使用，需要用比较简单且易于计算的函数 $P(x)$ 去近似代替它，使得

$$P(x_i)=y_i \qquad (i=0,1,2,\cdots,n) \tag{5.1}$$

这类问题称为插值问题。上述 $P(x)$ 称为插值函数，$x_0,x_1,\cdots,x_n$ 称为插值节点或简称为节点。这些插值节点所界的区间称为插值区间。条件(5.1)称为插值条件。从几何上看，插值函数就是通过 $(n+1)$ 个给定点 $(x_i,y_i)(i=0,1,2,\cdots,n)$ 的几何曲线。

插值函数为我们提供了一个重要的数学工具，利用它可以近似计算被插值函数 $f(x)$ 的函数值、极值、导数，并且用于数值积分、微分方程等数值解方面的近似计算。插值函数的形式可以是多项式、有理分式、三角函数、指数函数等。其中多项式或分段多项式最便于计算和使用，因而最引人注目，特别在计算机上被广泛地选作插值函数。本章只讨论多项式的插值问题，即构造 $n$ 次多项式

$$P_n(x)=a_0+a_1x+\cdots+a_nx^n \tag{5.2}$$

使满足

$$P_n(x_i)=y_i \qquad (i=0,1,2,\cdots,n) \tag{5.3}$$

及利用 $P_n(x)$ 进行插值计算的问题。

## §1　不等距节点下的牛顿基本差商公式

### 1.1　差商

对于可微函数，我们经常利用它的微商来研究函数的性质。同样，对待以表格形式给出的函数 $(x_i,y_i)(i=0,1,2,\cdots,n)$，我们经常利用的是它的差商。差商的定义如下：

$f(x)$ 在 $x_i$ 点的零阶差商为

$$f[x_i]=f(x_i) \qquad (i=0,1,2,\cdots,n) \tag{5.4}$$

$f(x)$ 在 $[x_i,x_j]$ 上的一阶差商为

$$f[x_i,x_j]=\frac{f[x_j]-f[x_i]}{x_j-x_i}=\frac{f(x_j)-f(x_i)}{x_j-x_i} \tag{5.5}$$

例如

$$f[x_0,x_1]=\frac{f(x_1)-f(x_0)}{x_1-x_0}$$

$$f[x_1,x_2]=\frac{f(x_2)-f(x_1)}{x_2-x_1}$$

$f(x)$ 在 $[x_i,x_j,x_k]$ 区间上一阶差商之差商为二阶差商

$$f[x_i,x_j,x_k]=\frac{f[x_j,x_k]-f[x_i,x_j]}{x_k-x_i} \tag{5.6}$$

例如 $[x_0,x_1,x_2]$ 区间上的二阶差商为

$$f[x_0,x_1,x_2]=\frac{f[x_1,x_2]-f[x_0,x_1]}{x_2-x_0}$$

$[x_1,x_2,x_3]$区间上的二阶差商为

$$f[x_1,x_2,x_3]=\frac{f[x_2,x_3]-f[x_1,x_2]}{x_3-x_1}$$

一般地,可定义区间$[x_i,x_{i+1},\cdots,x_{i+n}]$上的 $n$ 阶差商为

$$f[x_i,x_{i+1},\cdots,x_{i+n}]=\frac{f[x_{i+1},x_{i+2},\cdots,x_{i+n}]-f[x_i,x_{i+1},\cdots,x_{i+n-1}]}{x_{i+n}-x_i} \tag{5.7}$$

各阶差商可按表 5.1 的排列方式逐列进行计算,称表 5.1 为差商表。

**表 5.1**

| $x_i$ | $f[x_i]$ | $f[x_i,x_{i+1}]$ | $f[x_i,x_{i+1},x_{i+2}]$ | $f[x_i,x_{i+1},x_{i+2},x_{i+3}]$ |
|---|---|---|---|---|
| $x_0$ | $f(x_0)$ | | | |
| | | $f[x_0,x_1]$ | | |
| $x_1$ | $f(x_1)$ | | $f[x_0,x_1,x_2]$ | |
| | | $f[x_1,x_2]$ | | $f[x_0,x_1,x_2,x_3]$ |
| $x_2$ | $f(x_2)$ | | $f[x_1,x_2,x_3]$ | … |
| | | $f[x_2,x_3]$ | … | |
| $x_3$ | $f(x_3)$ | | | |
| | | … | | |
| … | … | | | |

**例 5.1** 试列出 $f(x)=x^3$ 在节点 $x=0,2,3,5,6$ 上的各阶差商值。

**解** 按表 5.1 计算得表 5.2。

**表 5.2**

| $x_i$ | $f[x_i]$ | $f[x_i,x_{i+1}]$ | $f[x_i,x_{i+1},x_{i+2}]$ | $f[x_i,x_{i+1},x_{i+2},x_{i+3}]$ |
|---|---|---|---|---|
| 0 | 0 | | | |
| | | $\frac{8-0}{2-0}=4$ | | |
| 2 | 8 | | $\frac{19-4}{3-0}=5$ | |
| | | $\frac{27-8}{3-2}=19$ | | $\frac{10-5}{5-0}=1$ |
| 3 | 27 | | $\frac{49-19}{5-2}=10$ | |
| | | $\frac{125-27}{5-3}=49$ | | $\frac{14-10}{6-2}=1$ |
| 5 | 125 | | $\frac{91-49}{6-3}=14$ | |
| | | $\frac{216-125}{6-5}=91$ | | |
| 6 | 216 | | | |

如以 $x$ 代表时间 $t$,$f(x)$代表路程 $s$,则一阶差商为 $\Delta s_i/\Delta t_i=\bar{v}_i$,它相当于在$[t_i,t_{i+1}]$范围内的一种平均速度,二阶差商则为上述平均速度的平均变化率,即平均加速度。所以差商表的数值可以直接反映出函数值的变化情况。差商具有一个重要的特性——对称性,即差商的值与同组节点排列的次序无关。如

$$f[x_0,x_1]=\frac{f[x_1]-f[x_0]}{x_1-x_0}=\frac{f[x_0]-f[x_1]}{x_0-x_1}=f[x_1,x_0] \tag{5.8}$$

$$f[x_0,x_1,x_2]=\frac{f[x_1,x_2]-f[x_0,x_1]}{x_2-x_0}$$
$$=\frac{1}{x_2-x_0}\left[\frac{f(x_2)-f(x_1)}{x_2-x_1}-\frac{f(x_1)-f(x_0)}{x_1-x_0}\right]$$
$$=\frac{f(x_0)}{(x_0-x_1)(x_0-x_2)}+\frac{f(x_1)}{(x_1-x_0)(x_1-x_2)}+\frac{f(x_2)}{(x_2-x_0)(x_2-x_1)} \tag{5.9}$$

若在式(5.9)左、右两边将$[x_0,x_1,x_2]$中的变量对应地更换为$[x_2,x_1,x_0]$或$[x_0,x_2,x_1]$，则得

$$\begin{cases} f[x_2,x_1,x_0]=\dfrac{f(x_2)}{(x_2-x_0)(x_2-x_1)}+\dfrac{f(x_1)}{(x_1-x_0)(x_1-x_2)}+\dfrac{f(x_0)}{(x_0-x_1)(x_0-x_2)} \\ f[x_0,x_2,x_1]=\dfrac{f(x_0)}{(x_0-x_1)(x_0-x_2)}+\dfrac{f(x_2)}{(x_2-x_0)(x_2-x_1)}+\dfrac{f(x_1)}{(x_1-x_0)(x_1-x_2)} \end{cases} \tag{5.10}$$

显见，它们的右端都是节点 $x_0,x_1,x_2$ 的对称函数，其差商值均相同。

一般情况下，可用数学归纳法证明 $f(x)$的 $k$ 阶差商 $f[x_0,x_1,\cdots,x_k]$是节点 $x_0,x_1,\cdots,x_k$ 的对称函数，并可表为如下的分解式

$$f[x_0,x_1,\cdots,x_k]=\frac{f(x_0)}{(x_0-x_1)(x_0-x_2)\cdots(x_0-x_k)}+\frac{f(x_1)}{(x_1-x_0)(x_1-x_2)\cdots(x_1-x_k)}+$$
$$\cdots+\frac{f(x_i)}{(x_i-x_0)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_k)}+\cdots+$$
$$\frac{f(x_k)}{(x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})} \tag{5.11}$$

因此不论如何安置 $x_0,x_1,\cdots,x_k$ 的排列顺序，它们所对应的差商值不变。

由表 5.2 可见，$f(x)=x^3$ 的三阶差商是一个常量。一般地，当 $f(x)=P_n(x)$为 $n$ 次多项式时，可以证明它的 $n$ 阶差商是一个常量(见本节 1.3)。

### 1.2 牛顿基本差商公式的建立

设 $x$ 为插值区间内的一个节点，按差商定义，有如下关系式

$$f[x_0,x]=\frac{f(x)-f(x_0)}{x-x_0} \tag{5.12}$$

$$f[x_1,x_0,x]=\frac{f[x_0,x]-f[x_1,x_0]}{x-x_1} \tag{5.13}$$

$$f[x_2,x_1,x_0,x]=\frac{f[x_1,x_0,x]-f[x_2,x_1,x_0]}{x-x_2} \tag{5.14}$$

$$\cdots$$

由式(5.12)、式(5.13)、式(5.14)…逐次解出 $f(x)$，$f[x_0,x]$，$f[x_1,x_0,x]$，$f[x_2,x_1,x_0,x]\cdots$ 并逐次代入下式得

$$f(x)=f(x_0)+(x-x_0)f[x_0,x]$$
$$=f(x_0)+(x-x_0)\{f[x_1,x_0]+(x-x_1)f[x_1,x_0,x]\}$$
$$=f(x_0)+(x-x_0)f[x_1,x_0]+(x-x_0)(x-x_1)\{f[x_2,x_1,x_0]+(x-x_2)f[x_2,x_1,x_0,x]\}$$
$$=f(x_0)+(x-x_0)f[x_1,x_0]+(x-x_0)(x-x_1)f[x_2,x_1,x_0]+(x-x_0)(x-x_1)(x-x_2)f[x_3,x_2,x_1,x_0]+\cdots+$$

$$(x-x_0)(x-x_1)\cdots(x-x_{n-1})f[x_n,x_{n-1},\cdots,x_1,x_0]+$$
$$(x-x_0)(x-x_1)\cdots(x-x_n)f[x_n,x_{n-1},\cdots,x_0,x]$$
$$=P_n(x)+R_n(x) \tag{5.15}$$

其中

$$P_n(x)=f(x_0)+(x-x_0)f[x_0,x_1]+(x-x_0)(x-x_1)f[x_0,x_1,x_2]+$$
$$(x-x_0)(x-x_1)(x-x_2)f[x_0,x_1,x_2,x_3]+\cdots+$$
$$(x-x_0)(x-x_1)\cdots(x-x_{n-1})f[x_0,x_1,\cdots,x_n] \tag{5.16}$$

称为牛顿基本差商公式。而

$$R_n(x)=(x-x_0)(x-x_1)\cdots(x-x_n)f[x_0,x_1,\cdots,x_n,x] \tag{5.17}$$

称为牛顿基本差商公式的余式。

由式(5.16)可见，牛顿基本差商公式是按函数 $f(x)$ 的各阶差商进行展开的，这些差商位于表 5.1 中的第一斜行上。若用 $P_n(x)$ 近似 $f(x)$，其误差为

$$R_n(x)=f(x)-P_n(x)=(x-x_0)(x-x_1)\cdots(x-x_n)f[x_0,x_1,\cdots,x_n,x] \tag{5.18}$$

当 $x=x_i(i=0,1,2,\cdots,n)$ 时，$R_n(x_i)=0$，即 $P_n(x_i)=y_i,i=0,1,2,\cdots,n$；而当 $x\neq x_i$ 时，一般有 $R_n(x)\neq 0$，所以 $P_n(x)\approx f(x)$。

**例 5.2** 已知 $x=1,4,9$ 的平方根值为 1,2,3，利用牛顿基本差商公式求 $\sqrt{7}$ 的近似值。

**解** 建立差商表 5.3。

表 5.3

| $x$ | $\sqrt{x}$ | $f[x_i,x_{i+1}]$ | $f[x_i,x_{i+1},x_{i+2}]$ |
|---|---|---|---|
| 1 | 1 | | |
| | | 0.333 33 | |
| 4 | 2 | | −0.016 67 |
| | | 0.2 | |
| 9 | 3 | | |

按式(5.16)得

$$P_2(7)=1+(7-1)\times 0.333\ 33+(7-1)(7-4)\times(-0.016\ 67)=2.699\ 92$$

## 1.3 牛顿基本差商公式的余式估计

### 1.3.1 差商与导数的关系式

考虑牛顿基本差商公式的余式 $R_n(x)=f(x)-P_n(x)$，当 $f(x)$ 为 $n$ 阶可导时，对余式 $n$ 阶求导有

$$R^{(n)}(x)=f^{(n)}(x)-P_n^{(n)}(x)$$
$$=f^{(n)}(x)-\{f(x_0)+(x-x_0)f[x_0,x_1]+\cdots+$$
$$(x-x_0)(x-x_1)\cdots(x-x_{n-1})f[x_0,x_1,\cdots,x_n]\}^{(n)}$$
$$=f^{(n)}(x)-n!f[x_0,x_1,\cdots,x_n] \tag{5.19}$$

设 $J$ 为由 $(n+1)$ 个节点 $x_0,x_1,\cdots,x_n$ 中最小值与最大值所确定的插值区间，则 $R_n(x)$ 在 $J$ 上具有 $(n+1)$ 个零点。由洛尔定理知 $R'_n(x)$ 在 $J$ 上必有 $n$ 个零点，$R''_n(x)$ 在 $J$ 上必有 $(n-1)$ 个零点…$R_n^{(n)}(x)$ 在 $J$ 上必有一个零点（设为 $\xi$），即

$$R_n^{(n)}(\xi) = 0, \qquad \xi \in J$$

成立。于是由式(5.19)推得差商与导数的关系式为

$$f[x_0, x_1, \cdots, x_n] = \frac{f^{(n)}(\xi)}{n!}, \qquad \xi \in J \tag{5.20}$$

若 $f(x)=P_n(x)$，因 $P_n^{(n)}(x)=$常量，因此任何 $n$ 次多项式的 $n$ 阶差商为一常量，其$(n+1)$阶差商为 0。若增加新节点 $x$ 且 $f(x)$为$(n+1)$阶可导时，同样有

$$f[x_0, x_1, \cdots, x_n, x] = \frac{f^{(n+1)}(\xi)}{(n+1)!}, \qquad \xi \in I \tag{5.21}$$

式中，$I$ 为由节点 $x_0, x_1, \cdots, x_n, x$ 所界的区间。

当差商公式中的变量相同时，按差商的狭义定义，其值为不定式 0/0。在这种情况下，应按极限的方法广义定义差商的值，对式(5.20)两边取极限得

$$\lim_{x_1, x_2, \cdots, x_n \to x_0} f[x_0, x_1, \cdots, x_n] = \lim_{\xi \to x_0} \frac{f^{(n)}(\xi)}{n!}$$

得

$$f[\underbrace{x_0, x_0, \cdots, x_0}_{(n+1)\text{个}}] = \frac{f^{(n)}(x_0)}{n!} \tag{5.22}$$

称式(5.22)为等变元的差商。当 $x_1, x_2, \cdots, x_n \to x_0$ 时，由(5.22)式知，这时牛顿基本差商公式转变为 $f(x)$在 $x_0$ 点的台劳展开式。显见，牛顿基本差商公式实际上就是台劳展开式的离散化表示。

### 1.3.2 余式 $R_n(x)$的估计

将(5.21)式代入(5.17)式得

$$R_n(x) = (x - x_0)(x - x_1)\cdots(x - x_n)\frac{f^{(n+1)}(\xi)}{(n+1)!} = \frac{f^{(n+1)}(\xi)}{(n+1)!}\prod\nolimits_{n+1}(x), \qquad \xi \in I \tag{5.23}$$

式中，$\prod_{n+1}(x)=(x-x_0)(x-x_1)\cdots(x-x_n)$。若在 $I$ 上有$|f^{(n+1)}(\xi)|\leqslant M_{n+1}$，则可得余式的估计式

$$|R_n(x)| \leqslant \frac{M_{n+1}}{(n+1)!}\left|\prod\nolimits_{n+1}(x)\right| \tag{5.24}$$

当 $x_1, x_2, \cdots, x_n \longrightarrow x_0$ 时，式(5.23)成为

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)^{n+1}, \qquad \xi \in (x_0, x) \tag{5.25}$$

它就是台劳公式的余式。

式(5.24)中的$|\Pi_{n+1}(x)|$的大小与节点 $x_0, x_1, \cdots, x_n$ 的分布情况有关。为了了解 $\Pi_{n+1}(x)$ 的特征，我们在图 5.1 和图 5.2 中画出了 $t^{[7]}=t(t-1)(t-2)\cdots(t-6)$和 $t^{[6]}=t(t-1)(t-2)\cdots(t-5)$的图形，该图形的特点是：①振幅两头大中间小。②当 $x$ 在插值区间以外时(称为外插)，其幅值甚大，所以外插误差较大，应当尽量避免之。

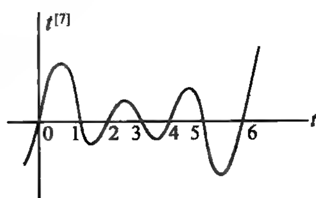



图 5.1

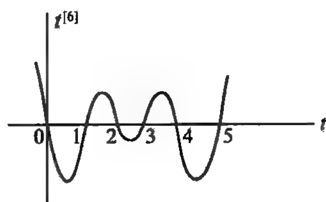



图 5.2

当节点为其他分布情况时，其变化特征相似。为使 $R_n(x)$ 较小，如果 $f(x)$ 是表格函数，这时节点只能从表格中选取，显然应选取距 $x$ 最邻近的 $n+1$ 个节点作为 $x_0,x_1,\cdots,x_n$，这样可使 $|\prod_{n+1}(x)|$ 较小。对解析函数来说，因 $|f^{(n+1)}(x)|$ 的最大值 $M_{n+1}$ 在插值区间内为确定值，如果选取 $x_0,x_1,\cdots,x_n$ 能使 $\max|\prod_{n+1}(x)|$ 达到最小，则余式的最大绝对值也为最小。为了达到这个目的，应选如下节点

$$x_k=\frac{b-a}{2}t_k+\frac{b+a}{2}\qquad(k=0,1,2,\cdots,n)\tag{5.26}$$

作为插值节点，其中 $t_k$ 为 $n+1$ 阶切比雪夫多项式 $\widetilde{T}_{n+1}(t)$ 的零点

$$t_k=\cos\frac{2k+1}{2(n+1)}\pi\qquad(k=0,1,2,\cdots,n)\tag{5.27}$$

这时 $|\prod_{n+1}(x)|$ 有以下估计公式

$$\left|\prod\nolimits_{n+1}(x)\right|\leqslant\frac{1}{2^n}\left(\frac{b-a}{2}\right)^{n+1}\tag{5.28}$$

(参看第八章 5.3.3。)

实际计算中，亦可采用事后估计误差的方法来估算 $R_n(x)$ 的大小，这是一种利用计算结果进行间接估计的方法。记 $P_n(x)$ 为以节点 $x_0,x_1,x_2,\cdots,x_n$ 建立的插值公式，另取一个新节点为 $x_{n+1}$，记 $P_n^{(1)}(x)$ 为以 $x_1,x_2,\cdots,x_n,x_{n+1}$ 建立的插值公式，相应的余式为

$$R_n(x)=f(x)-P_n(x)=\frac{f^{(n+1)}(\xi_1)}{(n+1)!}(x-x_0)(x-x_1)\cdots(x-x_n)$$

$$R_n^{(1)}(x)=f(x)-P_n^{(1)}(x)=\frac{f^{(n+1)}(\xi_2)}{(n+1)!}(x-x_1)(x-x_2)\cdots(x-x_n)(x-x_{n+1})$$

若 $f^{(n+1)}(x)$ 在插值区间上变化不大时，则有

$$\frac{f(x)-P_n(x)}{f(x)-P_n^{(1)}(x)}\approx\frac{x-x_0}{x-x_{n+1}}$$

从而可得

$$R_n(x)=f(x)-P_n(x)\approx\frac{x-x_0}{x_0-x_{n+1}}[P_n(x)-P_n^{(1)}(x)]\tag{5.29}$$

使得式(5.29)能方便地给出余式的实际估计。

**例 5.3**　用插值方法求 $\sqrt{7}$ 的近似值。

**解**　作函数 $f(x)=\sqrt{x}$，取 $x_0=4,x_1=9,x_2=6.25,x_3=4.84$；$y_0=2,y_1=3,y_2=2.5,y_3=2.2$，建立差商表 5.4。

表 5.4

| $x$ | $f(x)$ | $f[x_i,x_{i+1}]$ | $f[x_i,x_{i+1},x_{i+2}]$ |
|---|---|---|---|
| 4 | 2 | | |
| | | 0.2 | |
| 9 | 3 | | −0.008 08 |
| | | 0.181 82 | |
| 6.25 | 2.5 | | −0.007 44 |
| | | 0.212 77 | |
| 4.84 | 2.2 | | |

按牛顿基本差商公式计算得

$$P_2(7)=2+(7-4)\times 0.2+(7-4)(7-9)\times(-0.008\ 08)=2.648\ 48$$

在区间[4,9]上，$f^{(3)}(x)=\frac{3}{8}\left(\frac{1}{\sqrt{x}}\right)^5\leqslant\frac{3}{8}\left(\frac{1}{\sqrt{4}}\right)^5=0.011\ 719=M_3$，则 $R_2(7)\leqslant\frac{0.011\ 719}{3!}|(7-4)(7-9)(7-6.25)|\approx 0.008\ 79$。

若采用事后估计误差方法，另取节点 $x_1,x_2,x_3$ 为插值节点，建立牛顿基本差商公式计算得

$$P_2^{(1)}(7)=3+(7-9)\times 0.181\ 82+(7-9)(7-6.25)\times(-0.007\ 44)=2.647\ 52$$

按事后估计误差公式(5.29)可得到

$$f(7)-P_2(7)\approx\frac{7-4}{4-4.84}(2.648\ 48-2.647\ 52)=-0.003\ 43$$

它与实际值$\sqrt{7}-P_2(7)=-0.002\ 74$ 相差无几。在本题中，舍入误差较小，余式近似为 $0.5\times 10^{-2}$，$P_2(7)$可舍入为 2.65。

# §2 等距节点下的牛顿基本差商公式及弗雷瑟图表法

## 2.1 差分

设函数 $y=f(x)$在等距节点 $x_0,x_1,\cdots,x_n$ 上的值为 $y_0,y_1,\cdots,y_n$，则称

$$\Delta y_{i-1}=y_i-y_{i-1}\qquad(i=1,2,\cdots,n)\tag{5.30}$$

为 $f(x)$在$[x_{i-1},x_i]$上的一阶差分。称

$$\Delta^2 y_{i-1}=\Delta y_i-\Delta y_{i-1}\qquad(i=1,2,\cdots,n-1)\tag{5.31}$$

为 $f(x)$在$[x_{i-1},x_{i+1}]$上的二阶差分。推至一般，称

$$\Delta^k y_{i-1}=\Delta^{k-1}y_i-\Delta^{k-1}y_{i-1}\qquad(i=1,2,\cdots,n-\overline{k-1})\tag{5.32}$$

为 $f(x)$在$[x_{i-1},x_{i+k-1}]$上的 $k$ 阶差分。与差商表类似，各阶差分可按差分表 5.5 逐列进行计算。

表 5.5

| $x$ | $y$ | $\Delta y$ | $\Delta^2 y$ | $\Delta^3 y$ | $\Delta^4 y$ |
|---|---|---|---|---|---|
| $x_0$ | $y_0$ | | | | |
| | | $\Delta y_0$ | | | |
| $x_1$ | $y_1$ | | $\Delta^2 y_0$ | | |
| | | $\Delta y_1$ | | $\Delta^3 y_0$ | |
| $x_2$ | $y_2$ | | $\Delta^2 y_1$ | | $\Delta^4 y_0$ |
| | | $\Delta y_2$ | | $\Delta^3 y_1$ | |
| $x_3$ | $y_3$ | | $\Delta^2 y_2$ | | |
| | | $\Delta y_3$ | | | |
| $x_4$ | $y_4$ | | | | |

函数 $f(x)$ 的各阶差分亦可按各节点上的函数值表为

$$\Delta y_0 = y_1 - y_0$$
$$\Delta y_1 = y_2 - y_1$$
$$\Delta y_2 = y_3 - y_2$$
$$\cdots$$
$$\Delta^2 y_0 = \Delta y_1 - \Delta y_0 = y_2 - 2y_1 + y_0$$
$$\Delta^2 y_1 = \Delta y_2 - \Delta y_1 = y_3 - 2y_2 + y_1$$
$$\Delta^2 y_2 = \Delta y_3 - \Delta y_2 = y_4 - 2y_3 + y_2$$
$$\cdots$$
$$\Delta^3 y_0 = \Delta^2 y_1 - \Delta^2 y_0 = y_3 - 3y_2 + 3y_1 - y_0$$
$$\Delta^3 y_1 = \Delta^2 y_2 - \Delta^2 y_1 = y_4 - 3y_3 + 3y_2 - y_1$$
$$\Delta^3 y_2 = \Delta^2 y_3 - \Delta^2 y_2 = y_5 - 3y_4 + 3y_3 - y_2$$
$$\cdots$$
$$\Delta^4 y_0 = \Delta^3 y_1 - \Delta^3 y_0 = y_4 - 4y_3 + 6y_2 - 4y_1 + y_0$$
$$\Delta^4 y_1 = \Delta^3 y_2 - \Delta^3 y_1 = y_5 - 4y_4 + 6y_3 - 4y_2 + y_1$$
$$\Delta^4 y_2 = \Delta^3 y_3 - \Delta^3 y_2 = y_6 - 4y_5 + 6y_4 - 4y_3 + y_2$$
$$\cdots$$

等等。上述各阶差分中函数值的系数正好等于 $(a-b)^r(r=1,2,3,\cdots)$ 展开式中的系数。

在等距(间距为 $h$)节点情况下,节点为 $x_i=x_0+ih(i=0,1,\cdots,n)$,这时差商可以用差分表为

$$f[x_0,x_1]=\frac{\Delta y_0}{1!h},\quad f[x_1,x_2]=\frac{\Delta y_1}{1!h},\quad f[x_2,x_3]=\frac{\Delta y_2}{1!h},\cdots$$
$$f[x_0,x_1,x_2]=\frac{\Delta^2 y_0}{2!h^2},\quad f[x_1,x_2,x_3]=\frac{\Delta^2 y_1}{2!h^2},\quad f[x_2,x_3,x_4]=\frac{\Delta^2 y_2}{2!h^2},\cdots$$
$$f[x_0,x_1,x_2,x_3]=\frac{\Delta^3 y_0}{3!h^3},\quad f[x_1,x_2,x_3,x_4]=\frac{\Delta^3 y_1}{3!h^3},\cdots$$
$$\cdots$$

一般可表为

$$f[x_i,x_{i+1},\cdots,x_{i+n}]=\frac{\Delta^n y_i}{n!h^n}\qquad(i=0,1,2,\cdots)\tag{5.33}$$

因 $n$ 次多项式 $P_n(x)$ 的 $n$ 阶差商是一个常量,所以由式(5.33)知,$n$ 次多项式的 $n$ 阶差分亦是一个常量,即 $\Delta^n P_n(x)=$常量。

**例 5.4** 计算 $f(x)=x^3$ 在等距节点 0,1,2,3,4 上的各阶差分值。

**解** 见表 5.6。

**表 5.6**

| $x$ | $x^3$ | $\Delta y$ | $\Delta^2 y$ | $\Delta^3 y$ | $\Delta^4 y$ |
|---|---|---|---|---|---|
| 0 | 0 | 1 | 6 | 6 | 0 |
| 1 | 1 | 7 | 12 | 6 | |
| 2 | 8 | 19 | 18 | | |
| 3 | 27 | 37 | | | |
| 4 | 64 | | | | |

由于表格数据本身含有观测误差，即使是解析函数也会有舍入误差。在逐列建立差分表的过程中，假如某个函数值 $y_i$ 具有误差 $\varepsilon$，则该误差对各阶差分的影响如表 5.7 所示。它是随阶的升高按二项式系数分布的，且随着阶的升高而剧增。因此，在高阶差分中含有较大的误差成分。

表 5.7

| $y$ | $\Delta y$ | $\Delta^2 y$ | $\Delta^3 y$ | $\Delta^4 y$ |
|---|---|---|---|---|
| $y_{i-3}$ | | | | |
| | $\Delta y_{i-3}$ | | | |
| $y_{i-2}$ | | $\Delta^2 y_{i-3}$ | | $\Delta^4 y_{i-4}+\varepsilon$ |
| | $\Delta y_{i-2}$ | | $\Delta^3 y_{i-3}+\varepsilon$ | |
| $y_{i-1}$ | | $\Delta^2 y_{i-2}+\varepsilon$ | | $\Delta^4 y_{i-3}-4\varepsilon$ |
| | $\Delta y_{i-1}+\varepsilon$ | | $\Delta^3 y_{i-2}-3\varepsilon$ | |
| $y_i+\varepsilon$ | | $\Delta^2 y_{i-1}-2\varepsilon$ | | $\Delta^4 y_{i-2}+6\varepsilon$ |
| | $\Delta y_i-\varepsilon$ | | $\Delta^3 y_{i-1}+3\varepsilon$ | |
| $y_{i+1}$ | | $\Delta^2 y_i+\varepsilon$ | | $\Delta^4 y_{i-1}-4\varepsilon$ |
| | $\Delta y_{i+1}$ | | $\Delta^3 y_i-\varepsilon$ | |
| $y_{i+2}$ | | $\Delta^2 y_{i+1}$ | | $\Delta^4 y_i+\varepsilon$ |
| | $\Delta y_{i+2}$ | | | |
| $y_{i+3}$ | | | | |

如果差分表中的函数值都有舍入误差，设它们的误差限为 $0.5\times10^{-n}$（$n$ 表示值的小数位数），则 $\Delta y$ 可能具有的绝对误差限为 $2\times(0.5\times10^{-n})$，$\Delta^2 y$ 为 $2^2\times(0.5\times10^{-n})$，一般 $k$ 阶差分可能具有的绝对误差限为

$$\varepsilon_k = 2^k\times(0.5\times10^{-n}) = 2^{k-1}\times10^{-n} \tag{5.34}$$

当 $k$ 阶差分及所有低于 $k$ 阶的差分，其数值均大于 $\varepsilon_k$，而高于 $k$ 阶的差分数值均小于 $\varepsilon_k$ 时，则应舍弃那些小于 $\varepsilon_k$ 的差分部分而保留差分表至 $k$ 阶差分为止，称这部分差分为差分表的正常部分。使用差分表时，只应使用它的正常部分。例如，有 $y=e^x$ 的差分表 5.8，其 $\varepsilon_0=0.000\,5$，$\varepsilon_1=0.001$，$\varepsilon_2=0.002$，$\varepsilon_3=0.004$，$\varepsilon_4=0.008$，显见 $|\Delta^4 y|<\varepsilon_4$，则表 5.8 应截取至三阶差分作为它的正常部分。

表 5.8

| $x$ | $y$ | $\Delta y$ | $\Delta^2 y$ | $\Delta^3 y$ | $\Delta^4 y$ | $\Delta^5 y$ |
|---|---|---|---|---|---|---|
| 3.70 | 40.447 | | | | | |
| | | 2.074 | | | | |
| 3.75 | 42.521 | | 0.106 | | | |
| | | 2.180 | | 0.006 | | |
| 3.80 | 44.701 | | 0.112 | | −0.001 | |
| | | 2.262 | | 0.005 | | 0.003 |
| 3.85 | 46.993 | | 0.117 | | 0.002 | |
| | | 2.409 | | 0.007 | | −0.003 |
| 3.90 | 49.402 | | 0.124 | | −0.001 | |
| | | 2.533 | | 0.006 | | |
| 3.95 | 51.935 | | 0.130 | | | |
| | | 2.663 | | | | |
| 4.00 | 54.598 | | | | | |

## 2.2 牛顿前插公式

取间距为 $h$ 的等距节点 $x_0,x_1,\cdots,x_n$，建立牛顿基本差商公式，其中各阶差商用差分表示为

$$\begin{cases} f[x_0,x_1]=\dfrac{\Delta y_0}{1!h} \\ f[x_0,x_1,x_2]=\dfrac{\Delta^2 y_0}{2!h^2} \\ f[x_0,x_1,x_2,x_3]=\dfrac{\Delta^3 y_0}{3!h^3} \\ \quad\cdots \end{cases} \tag{5.35}$$

这样牛顿基本差商公式可化为

$$P_n(x) = y_0 + \frac{\Delta y_0}{1!h}(x-x_0) + \frac{\Delta^2 y_0}{2!h^2}(x-x_0)(x-x_1) + \frac{\Delta^3 y_0}{3!h^3}(x-x_0)(x-x_1)(x-x_2) + \cdots + \frac{\Delta^n y_0}{n!h^n}(x-x_0)(x-x_1)\cdots(x-x_{n-1}) \tag{5.36}$$

称上式为牛顿前插公式。如设

$$t = \frac{x-x_0}{h}$$

则 $x - x_i = (x-x_0)-(x_i-x_0) = (t-i)h$ 。代入式(5.36)后进一步化为

$$P_n(x) = y_0 + \frac{t}{1!}\Delta y_0 + \frac{t(t-1)}{2!}\Delta^2 y_0 + \frac{t(t-1)(t-2)}{3!}\Delta^3 y_0 + \cdots + \frac{t(t-1)\cdots(t-\overline{n-1})}{n!}\Delta^n y_0$$

为简化公式的记法,引入二项式系数的记号 $C_t^i$ 定义为

$$C_t^i = \frac{t(t-1)(t-2)\cdots(t-\overline{i-1})}{i!} = \frac{t^{[i]}}{i!} \tag{5.37}$$

式中,$t$ 为任意实数,$i$ 为自然数。则牛顿前插公式可简记为

$$P_n(x) = y_0 + C_t^1\Delta y_0 + C_t^2\Delta^2 y_0 + \cdots + C_t^n\Delta^n y_0 \tag{5.38}$$

式(5.38)中使用差分表 5.5 的第一斜行上的各阶差分。其余式为

$$\begin{aligned} R_n(x) &= \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)(x-x_1)\cdots(x-x_n) \\ &= \frac{f^{(n+1)}(\xi)}{(n+1)!}h^{n+1}t(t-1)\cdots(t-n) \qquad (\xi \in I) \end{aligned} \tag{5.39}$$

## 2.3 牛顿后插公式

在节点等距的情况下,若以相反的节点顺序 $x_n, x_{n-1}, \cdots, x_1, x_0$ 建立牛顿基本差商公式得

$$\begin{aligned} P_n(x) = f(x_n) &+ (x-x_n)f[x_n,x_{n-1}] + \\ &(x-x_n)(x-x_{n-1})f[x_n,x_{n-1},x_{n-2}] + \cdots + \\ &(x-x_n)(x-x_{n-1})\cdots(x-x_1)f[x_n,x_{n-1},\cdots,x_1,x_0] \end{aligned} \tag{5.40}$$

因为

$$\begin{cases} f[x_n,x_{n-1}] = f[x_{n-1},x_n] = \dfrac{\Delta y_{n-1}}{1!h} \\ f[x_n,x_{n-1},x_{n-2}] = f[x_{n-2},x_{n-1},x_n] = \dfrac{\Delta^2 y_{n-2}}{2!h^2} \\ f[x_n,x_{n-1},x_{n-2},x_{n-3}] = f[x_{n-3},x_{n-2},x_{n-1},x_n] = \dfrac{\Delta^3 y_{n-3}}{3!h^3} \\ \cdots \end{cases} \tag{5.41}$$

代入式(5.40)得

$$P_n(x) = y_n + \frac{\Delta y_{n-1}}{1!h}(x-x_n) + \frac{\Delta^2 y_{n-2}}{2!h^2}(x-x_n)(x-x_{n-1}) + \cdots + \frac{\Delta^n y_0}{n!h^n}(x-x_n)(x-x_{n-1})\cdots(x-x_1) \tag{5.42}$$

称上式为牛顿后插公式。为化简上式,引入变量

$$t = \frac{x - x_n}{h}$$

则 $x - x_{n-i} = (x - x_n) + (x_n - x_{n-i}) = (t+i)h$，代入式(5.42)得

$$P_n(x) = y_n + \frac{t}{1!}\Delta y_{n-1} + \frac{t(t+1)}{2!}\Delta^2 y_{n-2} + \frac{t(t+1)(t+2)}{3!}\Delta^3 y_{n-3} + \cdots + \frac{t(t+1)(t+2)\cdots(t+n-1)}{n!}\Delta^n y_0$$

$$= y_n + C_t^1 \Delta y_{n-1} + C_{t+1}^2 \Delta^2 y_{n-2} + C_{t+2}^3 \Delta^3 y_{n-3} + \cdots + C_{t+n-1}^n \Delta^n y_0 \tag{5.43}$$

式(5.43)中使用差分表 5.5 的最末斜行上的各阶差分。其余式为

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_n)(x - x_{n-1})\cdots(x - x_0)$$

$$= \frac{f^{(n+1)}(\xi)}{(n+1)!}h^{n+1}t(t+1)\cdots(t+n) \qquad \xi \in I \tag{5.44}$$

**例 5.5**　按下列数值表求 $y(-0.5)$和 $y(1.5)$的近似值。

| $x$ | $-1$ | 0 | 1 | 2 |
|---|---|---|---|---|
| $y$ | $-1$ | 1 | 3 | 11 |

**解**　建立差分表 5.9。

**表 5.9**

| $x$ | $y$ | $\Delta y$ | $\Delta^2 y$ | $\Delta^3 y$ |
|---|---|---|---|---|
| $-1$ | $-1$ | 2 | | |
| 0 | 1 | 2 | 0 | |
| 1 | 3 | 8 | 6 | 6 |
| 2 | 11 | | | |

后插公式

前插公式

① 按牛顿前插公式计算 $y(-0.5)$的近似值如下

$$t = \frac{(-0.5) - (-1)}{1} = 0.5$$

$$y(-0.5) \approx (-1) + \frac{0.5}{1!} \times 2 + \frac{0.5(0.5-1)}{2!} \times 0 + \frac{0.5(0.5-1)(0.5-2)}{3!} \times 6 = 0.375$$

② 按牛顿后插公式计算 $y(1.5)$的近似值如下

$$t = \frac{1.5-2}{1} = -0.5$$

$$y(1.5) \approx 11 + \frac{-0.5}{1!} \times 8 + \frac{-0.5(-0.5+1)}{2!} \times 6 + \frac{-0.5(-0.5+1)(-0.5+2)}{3!} \times 6 = 5.875$$

## 2.4 弗雷瑟(Fraser)图表及使用方法

下面叙述由弗雷瑟提议的方法，这是一个在等距节点下建立插值公式的一般方法。

### 2.4.1 弗雷瑟图表的建立方法及使用规则

以下经常使用二项式系数的记号 $C_t^m$，定义为

$$C_t^m = \frac{t^{[m]}}{m!} = \frac{t(t-1)(t-2)\cdots(t-m+1)}{m!}$$

其中 $m$ 为自然数，$t$ 为任意实数。它具有性质

$$C_{p+1}^{k+1} - C_p^{k+1} = C_p^k \tag{5.45}$$

由差分定义知

$$\Delta^{k+1} y_{s-1} = \Delta^k y_s - \Delta^k y_{s-1} \tag{5.46}$$

将式(5.45)与式(5.46)两边分别相乘得

$$C_{p+1}^{k+1}\Delta^{k+1} y_{s-1} - C_p^{k+1}\Delta^{k+1} y_{s-1} = C_p^k\Delta^k y_s - C_p^k\Delta^k y_{s-1}$$

或写成以下恒等式

$$C_p^k\Delta^k y_{s-1} + C_{p+1}^{k+1}\Delta^{k+1} y_{s-1} = C_p^k\Delta^k y_s + C_p^{k+1}\Delta^{k+1} y_{s-1} \tag{5.47}$$

如果把上式中的差分与二项式系数的乘积称为“项”，则式(5.47)是由四个项组成的恒等式，这种数量间的关系可在差分表中的差分间添加连结线与二项式系数构成的菱形来表达，如图5.3所示。

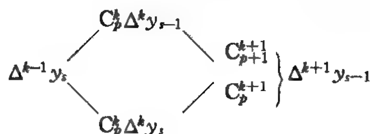



图 5.3

若从菱形的左顶点 $\Delta^{k-1} y_s$ 出发，分别沿菱形的上半边与下半边到达右顶点 $\Delta^{k+1} y_{s-1}$，则在上述两条路径上所得项之和分别为

$$I_1 = C_p^k\Delta^k y_{s-1} + C_{p+1}^{k+1}\Delta^{k+1} y_{s-1}$$
$$I_2 = C_p^k\Delta^k y_s + C_p^{k+1}\Delta^{k+1} y_{s-1}$$

由式(5.47)知 $I_1 = I_2$。

易于发现，若从 $\Delta^{k-1} y_s$ 出发，循菱形的水平对角线到达 $\Delta^{k+1} y_{s-1}$，并规定取该对角线上、下两个项的半和作为该路径上项之和，则有

$$I_3 = \frac{1}{2}(I_1 + I_2) = I_1 = I_2$$

成立。综上可知，由菱形的左顶点出发，沿三个不同方向(正斜率、负斜率、水平)的路径由左向右地到达菱形的右顶点，则它们所得的项之和都是相同的。

式(5.47)中的 $p$ 值目前是一个待定的参数，它的值可据差分与二项式系数间存在的特定关系确定。例如取 $s=1$ 和 $k=0$ 时，按式(5.47)得

$$y_0 + C_{p+1}^1\Delta y_0 = y_1 + C_p^1\Delta y_0$$

和以下存在的关系式

$$y_0 + C_t^1\Delta y_0 = y_1 + C_{t-1}^1\Delta y_0 \qquad \left(t = \frac{x - x_0}{h}\right)$$

对照知 $p=t-1$。一般当 $s=i$ 时，则有 $p=t-i$。当 $p$ 值这样确定后，关系式(5.47)化为

$$C_{t-i}^{k}\Delta^{k} y_{i-1} + C_{t-i+1}^{k+1}\Delta^{k+1} y_{i-1} = C_{t-i}^{k}\Delta^{k} y_i + C_{t-i}^{k+1}\Delta^{k+1} y_{i-1} \tag{5.48}$$

与图 5.3 类似，式(5.48)中的关系可用图 5.4 表达。在图 5.4 的基础上，继续拓展菱形图的范围，就可形成图 5.5，此图称为弗雷瑟图表或菱形图，它实际上就是在原差分表中添加所示的连结线与二项式系数构成的。为简化图 5.5，可将每个菱形的上下两个项点上相同的二项式系数合并成一个后放在菱形的中心位置上，如图 5.6 所示。

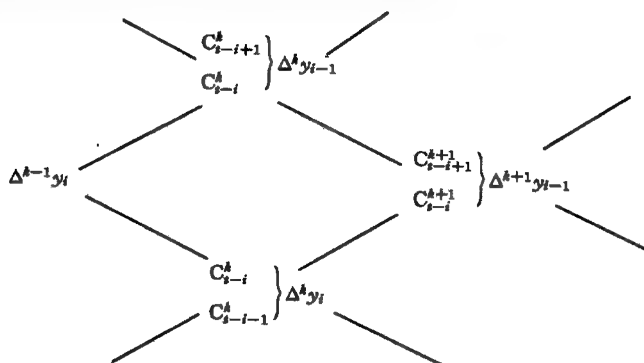



图 5.4

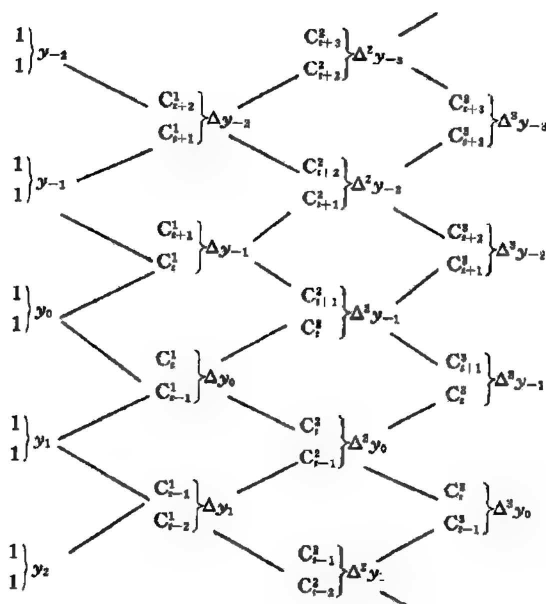



图 5.5

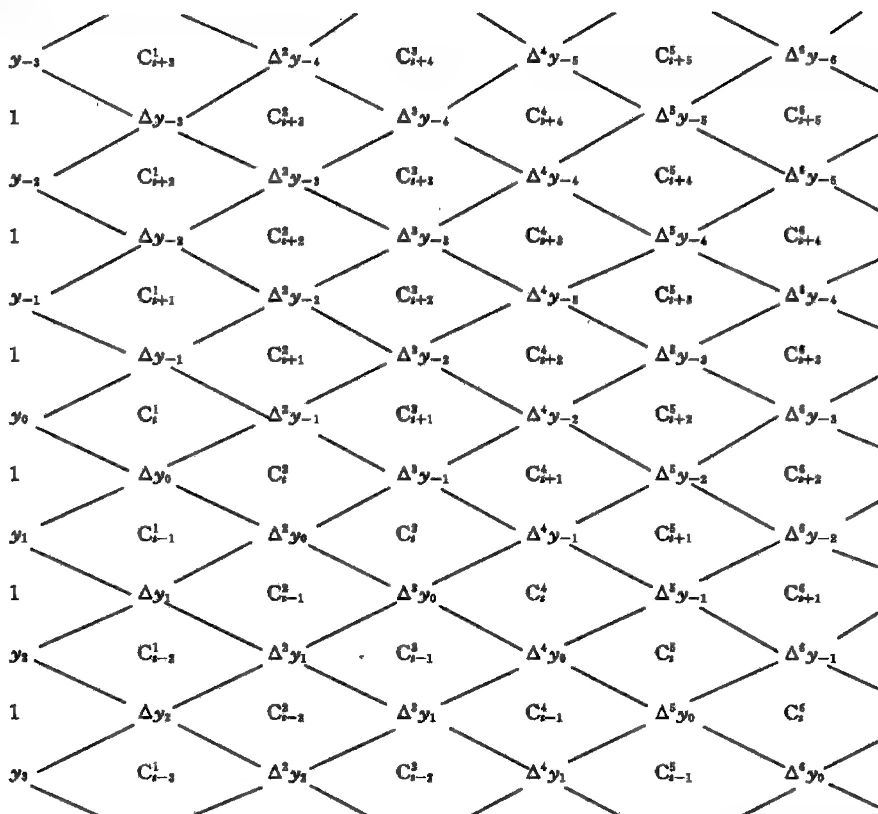


图 5.6

菱形图 5.6 中的项，可依据连结线的走向按以下规则来计算。

**规则 1** 若连结线依正斜率自左向右地相交于差分，则应取该差分与位于其下面的二项式系数之积作为项。

**规则 2** 若连结线依负斜率自左向右地相交于差分，则应取该差分与位于其上面的二项式系数之积作为项。

**规则 3** 若水平连结线自左向右地相交于二项式系数，则应取该二项式系数与位于其上下的二个差分的算术平均值之积作为项。

**规则 4** 若水平连结线自左向右地相交于差分，则应取该差分与位于其上下的二个二项式系数的算术平均值之积作为项。

**规则 5** 对每一条从右到左的连结线，可按该线段从左到右地相交于差分的规则 1～4 计算项后乘以－1。

利用以上规则 1～5，按照事先确定的路径，就可以使用菱形图 5.6 写出插值公式。这里所述的路径，就是从菱形图上的第一列的一个函数值(或相邻两个函数值的算术平均值)为出发点，沿着菱形的边或水平对角线前进，最终到达于某个差分(或上下两个差分的算术平均值)为止的一条折线。当路径前进至某一差分时，允许沿该差分周围的四条线段或水平对角线中的任一条继续前进，最后终止于一个差分上。使用图 5.6 建立插值公式的过程是：对上述任意选定的一条路径，写下路径出发时的函数值作为公式的第一项，然后在沿路径前进的过程中对每一线段按规则 1～5 给公式附加一个项，直至达到路径上最后一个差分所附加的项为止。

### 2.4.2 菱形图的性质

下面用引理和定理来阐明菱形图 5.6 按规则 1～5 建立插值公式的有关性质。

**引理 1** 当路径为整个菱形回路或菱形的上半个回路或下半个回路时，按规则 1～5 建立的插值公式，其各项之和均等于 0。

**证** 今考察图 5.3 中的一个菱形，由前知，由 $\Delta^{k-1}y_s$ 出发，分别循菱形的上边、下边或水平对角线按规则 1～4 所得的项之和分别为 $I_1$、$I_2$ 和 $I_3$，且 $I_1=I_2=I_3$。当路径为整个菱形回路时，按规则 1～5 所得各项之和为

$$I_1 - I_2 = 0$$

当路径为菱形的上半个回路时，按规则 1～5 所得各项之和为

$$I_1 - I_3 = 0$$

而当路径为菱形的下半个回路时，按规则 1～5 所得的各项之和为

$$I_3 - I_2 = 0$$

从而引理 1 得证。

**引理 2** 在菱形图 5.6 中，沿任何闭路径环绕时，按规则 1～5 所得各项之和为 0。

**证** 在菱形图 5.6 中，设有某个闭回路 $C$ 构成的路径，其环绕方向如图 5.7 中的箭头所示。若对包含于回路 $C$ 内的全部菱形按图 5.7 中所示方向分别环绕一周后，按规则 1～5 所得各项之和与环绕 $C$ 一周所得各项之和相同，据引理 1 知，其全部项之和为 0，引理 2 得证。

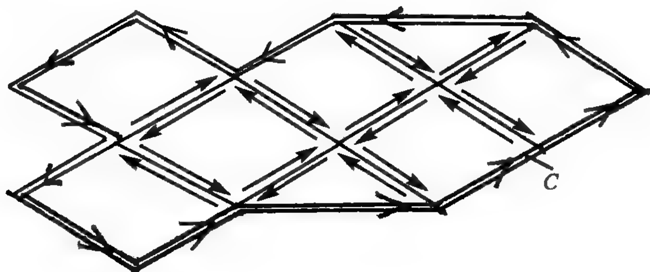



图 5.7

**定理 5.1** 在菱形图 5.6 中，对起于函数列的某个函数值 $y_i$（或 $(y_i+y_{i+1})/2$）和终止于同一个差分（或同列中上、下两个差分的算术平均值）的任何路径，按规则 1～5 建立的插值公式都是等价的（即它们具有相同的插值多项式）。

**证** 首先证明由 $y_0$ 出发，经过二条不同的路径到达于同一个差分，按规则 1～5 所建立的两个插值公式是等价的。设上述两个插值公式分别为 $y_0+s_1$，$y_0+s_2$，由于上述两条路径的起、末点相同，因此上述两条路径在菱形图中形成一个闭回路，据引理 2 知 $s_1=s_2$，证得上述两个插值公式是等价的。在此基础上，再证明经过 $y_0$ 及与其相邻的点 $(y_0+y_1)/2$、$y_1$ 分别沿不同路径到达于同一个差分的三个插值公式亦是彼此等价的。为此可设由 $y_0$ 出发，经过 $\Delta y_0$，再沿某一路径 $C_1$ 到达于某个差分的插值公式为

$$P_1 = y_0 + C_t^1 \Delta y_0 + s \tag{5.49}$$

另设 $C_2$ 为经过 $(y_0+y_1)/2$、$\Delta y_0$、$C_1$ 的路径；$C_3$ 为经过 $y_1$、$\Delta y_0$、$C_1$ 的路径，则相应于 $C_2$、$C_3$ 二条路径的插值公式为

$$P_2=\frac{1}{2}(y_0+y_1)+\frac{C_t^1+C_{t-1}^1}{2}\Delta y_0+s$$
$$=\frac{1}{2}(y_0+C_t^1\Delta y_0)+\frac{1}{2}[y_1+(t-1)\Delta y_0]+s$$
$$=\frac{1}{2}(y_0+C_t^1\Delta y_0)+\frac{1}{2}[(y_1-\Delta y_0)+t\Delta y_0]+s$$
$$=\frac{1}{2}(y_0+C_t^1\Delta y_0)+\frac{1}{2}(y_0+C_t^1\Delta y_0)+s=y_0+C_t^1\Delta y_0+s$$
$$P_3=y_1+C_{t-1}^1\Delta y_0+s$$
$$=y_0+\Delta y_0+(t-1)\Delta y_0+s$$
$$=y_0+C_t^1\Delta y_0+s \tag{5.50}$$

由此证得上述三个插值公式 $P_1$、$P_2$、$P_3$ 都是等价的。继而可知，经由 $y_0$、$(y_0+y_1)/2$、$y_1$ 而到达于同一个差分的任何路径所对应的插值公式亦都是等价的。按这种邻接关系递推，就可推知，只要终止的差分相同，则不同起点、不同路径所对应的插值公式都是等价的。（证毕）

这个定理可以用来建立等价或不等价的插值公式；同样可以用于验证不同插值公式间的等价或不等价性。虽然利用菱形图可以构造出多种多样的插值公式，按菱形图的性质可以证明，这样获得的公式等价于按终止的差分所涉及的节点上建立的插值多项式。

### 2.4.3 斯梯林插值公式和贝塞尔插值公式

利用菱形图，我们建立以下两个常用的插值公式。

(1) 斯梯林插值公式

在菱形图 5.6 中，若取过 $y_0$ 的水平线作为路径，按规则 1～5 可得

$$P_n(x)=y_0+C_t^1\frac{\Delta y_{-1}+\Delta y_0}{2}+\frac{C_{t+1}^2+C_t^2}{2}\Delta^2 y_{-1}+$$
$$C_{t+1}^3\frac{\Delta^3 y_{-2}+\Delta^3 y_{-1}}{2}+\frac{C_{t+2}^4+C_{t+1}^4}{2}\Delta^4 y_{-2}+\cdots+$$
$$C_{t+k-1}^{2k-1}\frac{\Delta^{2k-1}y_{-k}+\Delta^{2k-1}y_{-k+1}}{2}+\frac{C_{t+k}^{2k}+C_{t+k-1}^{2k}}{2}\Delta^{2k}y_{-k}+\cdots \tag{5.51}$$

上式称为斯梯林插值公式，式中所用的各阶差分对称分布于过 $x_0$ 的水平线，所需的插值节点同样对称于 $x_0$。各阶差分对应的系数计算如下

$$C_{t+k-1}^{2k-1}=\frac{1}{(2k-1)!}(t+k-1)^{[2k-1]}$$
$$=\frac{1}{(2k-1)!}[(t+\overline{k-1})(t+\overline{k-2})\cdots(t+1)t(t-1)\cdots(t-\overline{k-2})(t-\overline{k-1})$$
$$=\frac{1}{(2k-1)!}t(t^2-1)(t^2-2^2)\cdots(t^2-\overline{k-1}^{\,2}) \tag{5.52}$$

$$\frac{1}{2}[C_{t+k}^{2k}+C_{t+k-1}^{2k}]=\frac{1}{2(2k)!}[(t+k)^{[2k]}+(t+k-1)^{[2k]}]$$
$$=\frac{1}{2(2k)!}[(t+k)(t+k-1)^{[2k-1]}+(t+k-1)^{[2k-1]}(t+k-1-\overline{2k-1})]$$
$$=\frac{1}{2(2k)!}(t+k-1)^{[2k-1]}\cdot 2t \tag{5.53}$$
$$=\frac{1}{(2k)!}t^2(t^2-1)(t^2-2^2)\cdots(t^2-\overline{k-1}^{\,2})$$

代人式(5.51)后得

$$
\begin{aligned}
P_n(x) = & y_0 + \frac{t}{1!}\frac{\Delta y_{-1}+\Delta y_0}{2} + \frac{t^2}{2!}\Delta^2 y_{-1} + \\
& \frac{t(t^2-1)}{3!}\frac{\Delta^3 y_{-2}+\Delta^3 y_{-1}}{2} + \frac{t^2(t^2-1)}{4!}\Delta^4 y_{-2} + \\
& \frac{t(t^2-1)(t^2-2^2)}{5!}\frac{\Delta^5 y_{-3}+\Delta^5 y_{-2}}{2} + \frac{t^2(t^2-1)(t^2-2^2)}{6!}\Delta^6 y_{-3} + \cdots + \\
& \frac{t(t^2-1)(t^2-2^2)\cdots(t^2-\overline{k-1}^{\,2})}{(2k-1)!}\frac{\Delta^{2k-1} y_{-k}+\Delta^{2k-1} y_{-k+1}}{2} + \\
& \frac{t^2(t^2-1)(t^2-2^2)\cdots(t^2-\overline{k-1}^{\,2})}{(2k)!}\Delta^{2k} y_{-k} + \cdots
\end{aligned}
\tag{5.54}
$$

斯梯林插值公式的余式为

$$
\begin{aligned}
R_{2n}(x) &= \frac{1}{(2n+1)!}\frac{f^{(2n+1)}(\xi_1)+f^{(2n+1)}(\xi_2)}{2}h^{2n+1}t(t^2-1)(t^2-2^2)\cdots(t^2-n^2) \\
&= \frac{f^{(2n+1)}(\xi)}{(2n+1)!}h^{2n+1}t(t^2-1)(t^2-2^2)\cdots(t^2-n^2), \quad \xi \in I
\end{aligned}
\tag{5.55}
$$

$$
R_{2n-1}(x) = \frac{(t-n)f^{(2n)}(\xi_1)+(t+n)f^{(2n)}(\xi_2)}{2(2n)!}h^{2n}t(t^2-1)(t^2-2^2)\cdots(t^2-\overline{n-1}^{\,2}), \quad \xi_1,\xi_2 \in I
\tag{5.56}
$$

**例 5.6** 已知 $f(x)=\sin x$ 的差分表如表 5.10 所示。

**表 5.10**

| $x$ | $y$ | $\Delta y$ | $\Delta^2 y$ | $\Delta^3 y$ | $\Delta^4 y$ |
|---|---|---|---|---|---|
| 1.0 | 0.841 47 | | | | |
| | | 0.049 74 | | | |
| 1.1 | 0.891 21 | | −0.008 91 | | |
| | | 0.040 83 | | −0.000 40 | |
| 1.2 | 0.932 04 | | −0.009 31 | | 0.000 08 |
| | | 0.031 52 | | −0.000 32 | |
| 1.3 | 0.963 56 | | −0.009 63 | | 0.000 10 |
| | | 0.021 89 | | −0.000 22 | |
| 1.4 | 0.985 45 | | −0.009 85 | | 0.000 11 |
| | | 0.012 04 | | −0.000 11 | |
| 1.5 | 0.997 49 | | −0.009 96 | | 0.000 08 |
| | | 0.002 08 | | −0.000 03 | |
| 1.6 | 0.999 57 | | −0.009 99 | | |
| | | −0.007 91 | | | |
| 1.7 | 0.991 66 | | | | |

求 $y(1.31)$ 的近似值。

**解** 取 $x_0=1.3$，使用斯梯林插值公式计算如下

$$
t = \frac{1.31-1.3}{0.1} = 0.1
$$

$$
\begin{aligned}
y(1.31) \approx & 0.963\,56 + \frac{0.1}{1}\times\frac{0.031\,52+0.021\,89}{2} + \frac{0.1^2}{2}\times(-0.009\,63) + \\
& \frac{0.1(0.1^2-1)}{6}\times\frac{-0.000\,32+(-0.000\,22)}{2} + \frac{0.1^2(0.1^2-1)}{24}\times 0.000\,10 = 0.966\,19
\end{aligned}
$$

(2) 贝塞尔插值公式

在菱形图 5.6 中,若取过$\frac{y_0+y_1}{2}$的水平线作为路径,按规则 1~5 可得

$$P_n(x)=\frac{y_0+y_1}{2}+\frac{C_t^1+C_{t-1}^1}{2}\Delta y_0+C_t^2\frac{\Delta^2 y_{-1}+\Delta^2 y_0}{2}+\cdots+$$
$$\frac{C_{t+k-1}^{2k-1}+C_{t+k-2}^{2k-1}}{2}\Delta^{2k-1}y_{-k+1}+C_{t+k-1}^{2k}\frac{\Delta^{2k}y_{-k}+\Delta^{2k}y_{-k+1}}{2}+\cdots \tag{5.57}$$

称上式为贝塞尔插值公式,式中所用的各阶差分对称分布于过 $x_0$ 与 $x_1$ 之间中点$(x_0+x_1)/2$的水平线,所需的插值节点同样对称于中点$(x_0+x_1)/2$。各阶差分对应的系数计算如下

$$\frac{1}{2}[C_{t+k-1}^{2k-1}+C_{t+k-2}^{2k-1}]=\frac{1}{2(2k-1)!}[(t+k-1)^{[2k-1]}+(t+k-2)^{[2k-1]}]$$
$$=\frac{1}{2(2k-1)!}[(t+k-1)(t+k-2)^{[2k-2]}+(t+k-2)^{[2k-2]}(t+k-2-\overline{2k-2})]$$
$$=\frac{1}{2(2k-1)!}(t+k-2)^{[2k-2]}(2t-1)$$
$$=\frac{1}{2(2k-1)!}(t-\frac{1}{2})(t+k-2)^{[2k-2]} \tag{5.58}$$

$$C_{t+k-1}^{2k}=\frac{1}{(2k)!}(t+k-1)^{[2k]} \tag{5.59}$$

代入式(5.57)后得

$$P_n(x)=\frac{y_0+y_1}{2}+\frac{1}{1!}\left(t-\frac{1}{2}\right)\Delta y_0+\frac{t^{[2]}}{2!}\cdot\frac{\Delta^2 y_{-1}+\Delta^2 y_0}{2}+$$
$$\frac{1}{3!}\left(t-\frac{1}{2}\right)t^{[2]}\Delta^3 y_{-1}+\frac{(t+1)^{[4]}}{4!}\cdot\frac{\Delta^4 y_{-2}+\Delta^4 y_{-1}}{2}+$$
$$\frac{1}{5!}\left(t-\frac{1}{2}\right)(t+1)^{[4]}\Delta^5 y_{-2}+\frac{(t+2)^{[6]}}{6!}\cdot\frac{\Delta^6 y_{-3}+\Delta^6 y_{-2}}{2}+\cdots+$$
$$\frac{1}{(2k-1)!}\left(t-\frac{1}{2}\right)(t+k-2)^{[2k-2]}\Delta^{2k-1}y_{-k+1}+\frac{(t+k-1)^{[2k]}}{(2k)!}\cdot\frac{\Delta^{2k}y_{-k}+\Delta^{2k}y_{-k+1}}{2}+\cdots \tag{5.60}$$

当 $t=\frac{1}{2}$时,式(5.60)变得更加简单

$$P_n\left(\frac{x_0+x_1}{2}\right)=\frac{y_0+y_1}{2}-\frac{1}{8}\cdot\frac{\Delta^2 y_{-1}+\Delta^2 y_0}{2}+\frac{3}{128}\frac{\Delta^4 y_{-2}+\Delta^4 y_{-1}}{2}-$$
$$\frac{5}{1\,024}\frac{\Delta^6 y_{-3}+\Delta^6 y_{-2}}{2}+\cdots+(-1)^n\frac{[1\cdot 3\cdot 5\cdots(2n-1)]^2}{2^{2n}(2n)!}\cdot\frac{\Delta^{2n}y_{-n}+\Delta^{2n}y_{-n+1}}{2}+\cdots \tag{5.61}$$

这个公式叫做中点贝塞尔插值公式,可利用它加密表格值。

贝塞尔插值公式的余式为

$$R_{2n}(x)=\frac{h^{2n+1}}{2(2n+1)!}[f^{(2n+1)}(\xi_1)(t-\overline{n+1})+f^{(2n+1)}(\xi_2)(t+n)]t(t^2-1)(t^2-2^2)+\cdots+$$
$$(t^2-\overline{n-1}^{\,2})(t-n),\qquad \xi_1、\xi_2\in I \tag{5.62}$$

$$R_{2n-1}(x)=\frac{h^{2n}}{(2n)!}\frac{f^{(2n)}(\xi_1)+f^{(2n)}(\xi_2)}{2}t(t^2-1)(t^2-2^2)\cdots(t^2-\overline{n-1}^{\,2})(t-n)$$

$$= \frac{h^{2n}}{(2n)!} f^{(2n)}(\xi) t(t^2-1)(t^2-2^2)\cdots(t^2-\overline{n-1}^{\,2})(t-n), \qquad \xi_1、\xi_2、\xi \in I \tag{5.63}$$

**例 5.7** 利用例 5.6 的差分表计算 $y(1.33)$ 的近似值。

**解** 因为 $x=1.33$ 靠近 $x_0=1.3$ 与 $x_1=1.4$ 间的中线，应用贝塞尔插值公式计算如下

$$t=\frac{1.33-1.3}{0.1}=0.3$$

$$\begin{aligned} y(1.33) \approx & \frac{0.96356+0.98545}{2}+\frac{(0.3-0.5)}{1!}\times 0.02189+ \\ & \frac{0.3(0.3-1)}{2!}\times\frac{-0.00963+(-0.00985)}{2}+ \\ & \frac{(0.3-0.5)}{3!}\times 0.3(0.3-1)\times(-0.00022)+ \\ & \frac{(0.3+1)\times 0.3(0.3-1)(0.3-2)}{4!}\,\frac{0.00010+0.00011}{2} \\ = & 0.97115 \end{aligned}$$

斯梯林插值公式和贝塞尔插值公式都称为中心差分公式，它们都可用于 $x$ 位于插值区间中部插值计算用。

# §3 不等距节点下的拉格朗日插值公式

## 3.1 公式的建立

由差商的对称性知，$f[x_0,x_1,\cdots,x_n,x]$ 可按节点的函数值展开为

$$f[x_0,x_1,\cdots,x_n,x]=\frac{f(x)}{(x-x_0)(x-x_1)\cdots(x-x_n)}+\frac{f(x_0)}{(x_0-x_1)\cdots(x_0-x_n)(x_0-x)}+\cdots+\frac{f(x_n)}{(x_n-x_0)\cdots(x_n-x_{n-1})(x_n-x)} \tag{5.64}$$

在上式的两边用 $(x-x_0)(x-x_1)\cdots(x-x_n)$ 乘后再解出 $f(x)$ 得

$$\begin{aligned} f(x)= & \frac{(x-x_1)(x-x_2)\cdots(x-x_n)}{(x_0-x_1)(x_0-x_2)\cdots(x_0-x_n)}f(x_0)+\frac{(x-x_0)(x-x_2)\cdots(x-x_n)}{(x_1-x_0)(x_1-x_2)\cdots(x_1-x_n)}f(x_1)+\cdots+ \\ & \frac{(x-x_0)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)}{(x_i-x_0)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_n)}f(x_i)+\cdots+ \\ & \frac{(x-x_0)(x-x_1)\cdots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\cdots(x_n-x_{n-1})}f(x_n)+R_n(x) \\ = & L_n(x)+R_n(x) \end{aligned} \tag{5.65}$$

其中
$$R_n(x)=(x-x_0)(x-x_1)\cdots(x-x_n)f[x_0,x_1,\cdots,x_n,x]$$

称 $L_n(x)$ 为拉格朗日插值公式，$R_n(x)$ 为它的余式，这个余式与牛顿基本差商公式的余式完全相同。

## 3.2 拉格朗日插值公式的系数表达式

拉格朗日插值公式的应用极广，其系数表达的形式多样。诸如

$$a_i(x)=\frac{(x-x_0)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)}{(x_i-x_0)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_n)} \tag{5.66}$$

或

$$a_i(x)=\prod_{\substack{j=0\\j\neq i}}^{n}\left(\frac{x-x_j}{x_i-x_j}\right) \tag{5.67}$$

或

$$a_i(x)=\frac{\prod\limits_{\substack{j=0\\j\neq i}}^{n}(x-x_j)}{\prod\limits_{\substack{j=0\\j\neq i}}^{n}(x_i-x_j)} \tag{5.68}$$

或

$$a_i(x)=\frac{\prod_{n+1}(x)}{\prod'_{n+1}(x_i)(x-x_i)} \tag{5.69}$$

其中 $\prod_{n+1}(x)=(x-x_0)(x-x_1)\cdots(x-x_n)$。因

$$a_i(x)=\frac{(x-x_0)(x-x_1)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)}{(x_i-x_0)(x_i-x_1)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_n)}\cdot\frac{(x-x_i)}{(x-x_i)}$$

$$=\frac{\prod_{n+1}(x)}{[(x_i-x_0)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_n)](x-x_i)} \tag{5.70}$$

对 $\prod_{n+1}(x)$ 求导得

$$\begin{aligned}\prod\nolimits'_{n+1}(x)&=\{[(x-x_0)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)]\cdot(x-x_i)\}'\\&=(x-x_i)'[(x-x_0)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)]+\\&\quad(x-x_i)[(x-x_0)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)]'\\&=[(x-x_0)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)]+\\&\quad(x-x_i)[(x-x_0)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)]'\end{aligned}$$

令 $x=x_i$ 代入上式得

$$\prod\nolimits'_{n+1}(x_i)=(x_i-x_0)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_n) \tag{5.71}$$

代入式(5.70)知

$$a_i(x)=\frac{\prod_{n+1}(x)}{\prod'_{n+1}(x_i)(x-x_i)}\qquad(i=0,1,2,\cdots,n) \tag{5.72}$$

拉格朗日插值公式中 $f(x_i)$ 的系数 $a_i(x)$ 有如下特点:

$$a_i(x)=\begin{cases}0, & x=x_j\neq x_i\\1, & x=x_i\end{cases}\qquad(i,j=0,1,2,\cdots,n) \tag{5.73}$$

及

$$\sum_{i=0}^{n}a_i(x)\equiv 1 \tag{5.74}$$

因为 $f(x)\equiv L_n(x)+R_n(x)$，特别当 $f(x)=C$(常数)时，有

$$C\equiv\left[\sum_{i=0}^{n}a_i(x)\right]\cdot C$$

所以式(5.74)成立。式(5.74)常用来检查系数计算的正确性。称 $a_i(x)$ 为 $L_n(x)$ 的插值基函数。

### 3.3 拉格朗日插值公式舍入误差估计

设 $a_i^*(x)$、$y_i^*$、$L_n^*(x)$为精确值，$a_i(x)$、$y_i$、$L_n(x)$为其近似值，$\Delta a_i$、$\Delta y_i$、$\in$为其舍入误差，则有以下关系式

$$a_i^*(x) = a_i(x) + \Delta a_i, \quad y_i^* = y_i + \Delta y_i, \quad L_n^*(x) = L_n(x) + \in \tag{5.75}$$

代入

$$L_n^*(x) = \sum_{i=0}^{n} a_i^*(x) y_i^* \tag{5.76}$$

得

$$\begin{aligned} L_n(x) + \in &= \sum_{i=0}^{n} (a_i(x) + \Delta a_i)(y_i + \Delta y_i) \\ &= \sum_{i=0}^{n} a_i(x) y_i + \sum_{i=0}^{n} [\Delta y_i a_i(x) + \Delta a_i y_i + \Delta a_i \Delta y_i] \end{aligned}$$

可知近似值

$$L_n(x) = \sum_{i=0}^{n} a_i(x) y_i \tag{5.77}$$

的舍入误差(略去高阶量)为

$$\begin{aligned} |\in| &= \left| \sum_{i=0}^{n} [\Delta y_i \cdot a_i(x) + \Delta a_i \cdot y_i] \right| \\ &\leqslant \sum_{i=0}^{n} [|\Delta y_i| \cdot |a_i(x)| + |\Delta a_i| \cdot |y_i|] \end{aligned} \tag{5.78}$$

当$|\Delta y_i| = |\Delta a_i| = \Delta$时，上式化为

$$|\in| \leqslant \left( \sum_{i=0}^{n} [|a_i(x)| + |y_i|] \right) \cdot \Delta \tag{5.79}$$

当 $a_i(x)$为精确值时，$\Delta a_i = 0$，式(5.78)化为

$$|\in| \leqslant \left( \sum_{i=0}^{n} |a_i(x)| \right) \cdot |\Delta y| \tag{5.80}$$

其中$|\Delta y_0| = |\Delta y_1| = \cdots = |\Delta y_n| = |\Delta y|$。因为

$$\sum_{i=0}^{n} |a_i(x)| \begin{cases} = \sum_{i=0}^{n} a_i(x) = 1, \text{当所有 } a_i(x) > 0 \text{ 时} \\ > 1, \text{当部分 } a_i(x) > 0, \text{部分 } a_i(x) < 0 \text{ 时} \end{cases}$$

可见，当拉格朗日插值公式中有负系数出现时，它们对 $y_i$ 的舍入误差有放大作用。

**例 5.8** 试估计用线性插值法计算 lg47 时的误差限。使用表 5.11。

**表 5.11**

| $x$ | 42 | 45 | 48 |
|---|---|---|---|
| $\lg x$ | 1.623 249 3 | 1.653 212 6 | 1.681 241 3 |

**解** 应用 $n=1$ 时的拉格朗日插值公式

$$y = \frac{x - x_1}{x_0 - x_1} y_0 + \frac{x - x_0}{x_1 - x_0} y_1 \tag{5.81}$$

其中，取 $x_0 = 45$，$x_1 = 48$，按上式计算得

$$y=\frac{47-48}{45-48}y_0+\frac{47-45}{48-45}y_1$$
$$=0.3333333y_0+0.6666667y_1$$
$$=1.671898401$$

余式为 $$R_1(x)=\frac{1}{2}\lg''\xi(x-x_0)(x-x_1)\qquad(45<\xi<48)$$

估计 $\lg''\xi$ 的大小

$$(\lg x)'=\frac{\lg e}{x},\qquad(\lg x)''=-\frac{\lg e}{x^2}=-\frac{0.43}{x^2}$$

所以 $$|R_1(47)|\leqslant\left|\frac{1}{2}\times\frac{-0.43}{45^2}\times(47-45)(47-48)\right|=0.2\times10^{-3}$$

结果的舍入误差限为

$$\in=(0.3333333+1.6532126)\times(0.5\times10^{-7})+$$
$$(0.6666667+1.6812413)\times(0.5\times10^{-7})=0.2\times10^{-6}$$

结果的总误差为

$$\varepsilon=0.2\times10^{-3}+0.2\times10^{-6}\approx0.2\times10^{-3}<0.5\times10^{-3}$$

所以可取 $y=1.672$。

**例 5.9** 设 $y=\sin x$，当取用 $x_0=1.74, x_1=1.76, x_2=1.78$ 建立拉格朗日插值公式计算 $x=1.75$ 的函数近似值时，函数值 $y_0=\sin x_0, y_1=\sin x_1, y_2=\sin x_2$ 应取几位小数计算为好？

**解** 建立二次拉格朗日插值公式计算得

$$L_2(1.75)=0.375y_0+0.75y_1-0.125y_2\tag{5.82}$$

$$|R_2(1.75)|\leqslant\left|\frac{\sin^{(3)}\xi}{3!}(1.75-1.74)(1.75-1.76)(1.75-1.78)\right|$$
$$\leqslant\frac{1}{3!}|(1.75-1.74)(1.75-1.76)(1.75-1.78)|=0.5\times10^{-6}$$

$$|\in|\leqslant(0.375+0.75+0.125)\cdot|\Delta y|=1.25|\Delta y|\tag{5.83}$$

令 $|\in|=|R_2(1.75)|=0.5\times10^{-6}$ 及 $1.25|\Delta y|\leqslant0.5\times10^{-6}$ 得

$$|\Delta y|\leqslant\frac{0.5\times10^{-6}}{1.25}=0.0000004$$

因为 $0.00000005<0.0000004$，所以可取 $|\Delta y|=0.5\times10^{-7}$，即 $y_0, y_1, y_2$ 应取 7 位小数比较合宜。

# §4 等距节点下的拉格朗日插值公式

## 4.1 等距节点下的拉格朗日插值公式

对等距节点 $x_i=x_0+ih(i=0,1,2,\cdots,n)$ 引入变量

$$t=\frac{x-x_0}{h}$$

则有 $x-x_i=(t-i)h(i=0,1,2,\cdots,n)$。代入拉格朗日插值公式得

$$L_n(x)=\sum_{i=0}^{n}\frac{(x-x_0)(x-x_1)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)}{(x_i-x_0)(x_i-x_1)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_n)}f(x_i)$$

$$= \sum_{i=0}^{n} \frac{th \cdot (t-1)h \cdots (t-\overline{i-1})h \cdot (t-\overline{i+1})h \cdots (t-n)h}{ih \cdot (i-1)h \cdots 3h \cdot 2h \cdot 1h \cdot (-1h) \cdot (-2h) \cdots [-(n-i)h]} f(x_i)$$

$$= \sum_{i=0}^{n} \frac{t(t-1)(t-2)\cdots(t-\overline{i-1})(t-\overline{i+1})\cdots(t-n)}{[i(i-1)(i-2)\cdots 3 \cdot 2 \cdot 1] \cdot (-1)^{n-i}[1 \cdot 2 \cdot 3 \cdots (n-i)]} f(x_i)$$

$$= \sum_{i=0}^{n} \frac{[t(t-1)(t-2)\cdots(t-\overline{i-1})(t-\overline{i+1})\cdots(t-n)](t-i)}{(-1)^{n-i} i!(n-i)!(t-i)} f(x_i)$$

$$= \sum_{i=0}^{n} \frac{(-1)^{n-i} t^{[n+1]}}{i!(n-i)!(t-i)} f(x_i) \tag{5.84}$$

其中 $t \neq i, t^{[n+1]} = t(t-1)\cdots(t-n)$。

## 4.2 等距节点下的分段线性插值公式

当 $n=1$ 时,式(5.84)化为线性插值公式

$$\begin{cases} L_1(x) = (1-t)f(x_0) + tf(x_1) = f(x_0) + [f(x_1) - f(x_0)]t \\ t = \dfrac{x - x_0}{h}, \qquad x \in [x_0, x_1], t \in [0,1] \end{cases} \tag{5.85}$$

当 $x \in [x_i, x_{i+1}]$ 时,相应地有

$$\begin{cases} L_1(x) = f(x_i) + [f(x_{i+1}) - f(x_i)]t \\ t = \dfrac{x - x_i}{h}, \qquad t \in [0,1] \end{cases} \tag{5.86}$$

其余式为

$$|R_1(x)| = \left| \frac{f''(\xi)}{2!}(x-x_i)(x-x_{i+1}) \right| \leqslant \frac{M_2}{2} h^2 t(t-1) \Big|_{t=\frac{1}{2}} = \frac{M_2}{8} h^2 \tag{5.87}$$

其中 $|f''(x)| \leqslant M_2, \quad x \in [x_i, x_{i+1}] \quad (i = 0,1,2,\cdots,n-1)$

式(5.86)就是在等距节点下的分段线性插值公式。其间距 $h$ 可由对 $|R_1(x)|$ 的精度要求按式(5.87)确定。因在每一小段 $[x_i, x_{i+1}]$ 上用低次插值多项式来近似函数,其舍入误差和计算量均较小。在计算中要解决的问题是 $x \in [x_i, x_{i+1}]$ 的判断问题及确定 $t$ 值的大小问题。因 $t = \frac{x - x_i}{h}$,从下式

$$\frac{x - x_0}{h} = \frac{(x_i + th) - x_0}{h} = \frac{x_i - x_0}{h} + t = i + t \tag{5.88}$$

可知

$$i = \left[ \frac{x - x_0}{h} \right], \quad t = \left\{ \frac{x - x_0}{h} \right\} \tag{5.89}$$

式中,[]为取整值部分运算;{}为取小数部分运算。

特别是在计算机上,当 $x_0 = 0, h = 2^{-k}$($k$ 为正整数)时,式(5.89)成为

$$i = [2^k x], \quad t = \{2^k x\} \tag{5.90}$$

上式只需将 $x$ 左移 $k$ 位后截取其整数部分和小数部分就可得到 $i$ 与 $t$ 值,根据 $i$ 值取出 $y_i$、$y_{i+1}$ 的数值,与 $t$ 值一起代入式(5.86)就可进行插值计算。

## 4.3 等距节点下的分段三点插值公式

在等距节点下,对于给定的 $x$,设与 $x$ 最靠近的节点为 $x_k$,则 $x$ 满足下式

$$x_k-\frac{h}{2}<x<x_k+\frac{h}{2}$$

或

$$x_k<x+\frac{h}{2}<x_k+h=x_{k+1} \tag{5.91}$$

因此下标 $k$ 值为

$$k=\left[\frac{\left(x+\frac{h}{2}\right)-x_0}{h}\right]=\left[\frac{x-x_0}{h}+\frac{1}{2}\right] \tag{5.92}$$

据 $k$ 值就可取定以下三点值$(x_{k-1},y_{k-1})$，$(x_k,y_k)$，$(x_{k+1},y_{k+1})$。过上述三点建立二次拉格朗日插值公式

$$L_2(x)=\frac{(x-x_k)(x-x_{k+1})}{(x_{k-1}-x_k)(x_{k-1}-x_{k+1})}y_{k-1}+$$

$$\frac{(x-x_{k-1})(x-x_{k+1})}{(x_k-x_{k-1})(x_k-x_{k+1})}y_k+\frac{(x-x_{k-1})(x-x_k)}{(x_{k+1}-x_{k-1})(x_{k+1}-x_k)}y_{k+1} \tag{5.93}$$

令 $t=\frac{x-x_k}{h}$，则得

$$x-x_k=th,x-x_{k-1}=(x-x_k)+(x_k-x_{k-1})=(t+1)h,$$

$$x-x_{k+1}=(x-x_k)+(x_k-x_{k+1})=(t-1)h$$

代入式(5.93)后成为

$$\begin{cases}L_2(x)=\frac{t^2-t}{2}y_{k-1}+(1-t^2)y_k+\frac{t^2+t}{2}y_{k+1}\\ t=\frac{x-x_k}{h}\end{cases} \tag{5.94}$$

计算中需要特殊处理的情况是：当 $x\leqslant x_0+h/2$ 时，$k\leqslant 1$，这时应规定 $x_k=x_1$，取定 $x_0$、$x_1$、$x_2$ 三节点进行插值计算。另当 $x\geqslant x_{n-1}+h/2$ 时，$k\geqslant n$，这时应规定 $x_k=x_{n-1}$，取定 $x_{n-2}$、$x_{n-1}$、$x_n$ 三节点进行插值计算。

上述情况的另一种处理办法是，利用$(x_0,y_0)$、$(x_1,y_1)$、$(x_2,y_2)$的三点插值公式计算出 $y_{-1}=L_2(x_{-1})$；同法利用$(x_{n-2},y_{n-2})$、$(x_{n-1},y_{n-1})$、$(x_n,y_n)$的三点插值公式计算出 $y_{n+1}=L_2(x_{n+1})$。设插值节点为 $x_i=x_0+ih(i=-1,0,1,2,\cdots,n,n+1)$，对于 $x\in[x_0,x_n]$，可利用以下公式进行插值计算

$$L_2(x)=\frac{t^2-t}{2}y_{k-1}+(1-t^2)y_k+\frac{t^2+t}{2}y_{k+1} \tag{5.95}$$

其中

$$t=\frac{x-x_k}{h}$$

$$k=\left[\frac{x-x_0}{h}+\frac{1}{2}\right]$$

# §5　插值公式的唯一性及其应用

## 5.1　插值公式的唯一性

如果插值节点相同，在满足插值条件下用不同方法所建立的插值公式是相同(或等价)还

是不同(或不等价)的呢？下面证明插值公式的唯一性问题。

设 $P_n(x)$ 与 $Q_n(x)$ 为两个满足同一插值条件的不同的 $n$ 次插值多项式，则

$$G_n(x) = P_n(x) - Q_n(x) \tag{5.96}$$

为次数不超过 $n$ 的多项式，据插值条件知，当 $x=x_0,x_1,\cdots x_n$ 时，$G_n(x_i)=0(i=0,1,2,\cdots,n)$。即 $G_n(x)$ 具有 $(n+1)$ 个不同的零点，这是不可能的。这说明上述假设不对，从而证得 $P_n(x)=Q_n(x)$，即满足插值条件的次数不超过 $n$ 的多项式都是相同的。

## 5.2　插值公式的应用

在不等距节点情况下，若采用牛顿基本差商公式，则当发现公式精度不够需再增加一个新的节点时，只需在原来的结果上增添一项；而采用拉格朗日插值公式时，则都要重新计算。但在估算结果的舍入误差时，使用拉格朗日插值公式比较容易。

当节点为等距分布时，如果 $x$ 靠近表头，我们自然要挑选和 $x$ 最靠近的 $x_0,x_1,\cdots,x_n$ 作为插值节点，这时选用牛顿前插公式比较方便。同样理由，牛顿后插公式主要用在 $x$ 处在表末时插值计算用。如果 $x$ 处于插值区间中部，则可采用中心差分公式进行插值计算。它们的选用应考虑以下两方面。

① 从插值节点相对于 $x$ 的对称分布着眼，当 $x$ 靠近某插值节点(设为 $x_0$)时，即 $|t|\leqslant 1/4$ 时，以选用斯梯林插值公式为好。当 $x$ 靠近两节点(设为 $x_0,x_1$)间的中点时，即 $|t-1/2|\leqslant 1/4$ 时，以选用贝塞尔插值公式为好。

② 从余式简单，便于估算着眼，如果希望插值公式截止到偶阶差分时，以使用斯梯林插值公式为好。如果希望插值公式截止到奇阶差分时，以选用贝塞尔插值公式为好。

使用插值多项式来近似函数时，由余式可见，其逼近的程度不单与插值节点的个数及其分布情况有关，还与函数本身有关。直觉上，似乎插值节点愈密，相应的插值多项式的次数愈高，被插值函数与插值函数间的差别应愈小，实际情况却往往不是这样。对于具有各阶连续导数的光滑函数来说，随着节点数的增加，有的函数，例如整函数的插值多项式收敛于函数本身；有的就不一定收敛于该函数。德国数学家龙格(Runge)曾给出这样一个例子，取 $f(x)=\dfrac{1}{1+x^2}$，这是一个很光滑的函数，它的任意阶导数都存在。今在 $[-5,5]$ 区间上使用 $n+1$ 个节点 $x_i=-5+\dfrac{10}{n}i\,(i=0,1,2,\cdots,n)$ 作 $n$ 次插值多项式 $P_n(x)$，龙格已经证明，当 $n\to\infty$ 时，$P_n(x)$ 不在整个区间 $[-5,5]$ 上收敛于 $f(x)$。图 5.5 中绘出了 $y=f(x)$(实线表示)的曲线与 $y=P_{10}(x)$(虚线表示)的曲线，可以看出，在 $x=\pm 5$ 附近，$P_{10}(x)$ 偏离 $f(x)$ 很大，例如 $P_{10}(4.8)\approx 1.8$，$f(4.8)\approx 0.04$。这种插值多项式当节点增加时依然不能很好地逼近被插值函数的现象称为"龙格现象"。这个例子说明，虽然节点增多能使插值多项式在更多的节点上与 $f(x)$ 重合，但并不能保证在节点间插值多项式能很好地逼近 $f(x)$。龙格现象表明，为减少逼近误差，盲目地提高插值多项式的次数是不可取的；而且从计算的舍入误差来看，高次插值的误差累积可以泛滥成灾。从公式(5.20)可见，对于有间断导数的函数，差商与导数间差别较大，因此取插值多项式的次数小于具有间断的导数的阶数为宜。在实际应用中，高次插值(如七八次以上)极少被采用，而采用加密节点的分段低次多项式插值的方法往往能收到较好的效果。

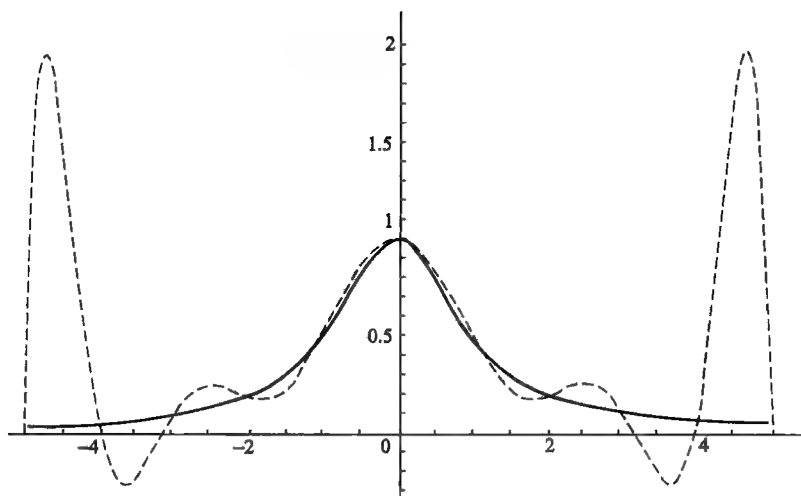



图 5.8

# §6 反 插 值

反插值问题是求对应于已知 $y$ 值的 $x$ 值，可以采用下面两种办法解决。

## 6.1 使用反函数的插值法

我们将 $x$ 视为 $y$ 的函数，即 $x=\varphi(y)$，因此可直接利用插值公式进行反插值计算。反函数的拉格朗日插值公式是

$$x=\sum_{i=0}^{n}\frac{(y-y_0)\cdots(y-y_{i-1})(y-y_{i+1})\cdots(y-y_n)}{(y_i-y_0)\cdots(y_i-y_{i-1})(y_i-y_{i+1})\cdots(y_i-y_n)}x_i$$

$$=\sum_{i=0}^{n}a_i(y)x_i \tag{5.97}$$

反函数的牛顿基本差商公式是

$$x=x_0+(y-y_0)\varphi[y_0,y_1]+(y-y_0)(y-y_1)\varphi[y_0,y_1,y_2]+\cdots+$$
$$(y-y_0)(y-y_1)\cdots(y-y_{n-1})\varphi[y_0,y_1,\cdots,y_n] \tag{5.98}$$

必须指出，这种将 $x$ 和 $y$ 关系反转的办法只在 $y=f(x)$ 为单调连续函数时才对，否则会出现同一 $y$ 值对应多个 $x$ 值的问题。

**例 5.10** 设有 8 位 $\sin x$ 的函数表如表 5.12 所示。

**表 5.12**

| $x$ | 1.74 | 1.76 | 1.78 | 1.80 |
|---|---|---|---|---|
| $\sin x$ | 0.985 719 18 | 0.982 154 32 | 0.978 196 61 | 0.973 847 63 |

对 $y=0.980\ 000\ 00$ 利用 $y=\sin x$ 的反函数进行反插值。

**解** 取下列数据，见表 5.13。

表 5.13

| $i$ | 0 | 1 | 2 | 3 |
| --- | --- | --- | --- | --- |
| $y_i$ | 0.985 719 18 | 0.982 154 32 | 0.978 196 61 | 0.973 847 63 |
| $x_i$ | 1.74 | 1.76 | 1.78 | 1.80 |

代入反插值的拉格朗日插值公式得

$$x=\frac{(0.98-y_1)(0.98-y_2)(0.98-y_3)}{(y_0-y_1)(y_0-y_2)(y_0-y_3)}\times 1.74+\frac{(0.98-y_0)(0.98-y_2)(0.98-y_3)}{(y_1-y_0)(y_1-y_2)(y_1-y_3)}\times 1.76+\frac{(0.98-y_0)(0.98-y_1)(0.98-y_3)}{(y_2-y_0)(y_2-y_1)(y_2-y_3)}\times 1.78+\frac{(0.98-y_0)(0.98-y_1)(0.98-y_2)}{(y_3-y_0)(y_3-y_1)(y_3-y_2)}\times 1.80$$

$$=-0.075\,080\,33\times 1.74+0.541\,441\,34\times 1.76+0.585\,448\,64\times 1.78-0.051\,809\,67\times 1.80$$

$$=1.77\,113\,820$$

如建立反函数的差商表 5.14，对 $y=0.98$，利用反函数的牛顿基本差商公式计算得

$$x=1.74+(0.98-0.985\,719\,18)\times(-5.610\,318\,50)+(0.98-0.985\,719\,18)\times(0.98-0.982\,154\,32)\times(-74.029\,372\,94)+(0.98-0.985\,719\,18)\times(0.98-0.982\,154\,32)\times(0.98-0.978\,196\,61)\times(-1\,625.452\,096)=1.771\,138\,19$$

表 5.14

| $y$ | $x$ | $\varphi[y_i,y_{i+1}]$ | $\varphi[y_i,y_{i+1},y_{i+2}]$ | $\varphi[y_i,y_{i+1},y_{i+2},y_{i+3}]$ |
| --- | --- | --- | --- | --- |
| 0.985 719 18 | 1.74 | | | |
| | | −5.610 318 50 | | |
| 0.982 154 32 | 1.76 | | −74.029 372 94 | |
| | | −5.053 427 36 | | −1 625.452 096 |
| 0.978 196 61 | 1.78 | | −54.732 737 11 | |
| | | −4.598 779 48 | | |
| 0.973 847 63 | 1.80 | | | |

## 6.2 利用正函数插值公式的反插值法

### 6.2.1 方法描述

考虑在 $x_0$ 与 $x_1$ 之间进行反插值。设 $c$ 是 $y_0$ 与 $y_1$ 之间的一个值，则当 $f(x)$ 连续时，在 $x_0$ 与 $x_1$ 之间必有 $x$，使 $f(x)=c$。今利用如下插值多项式来进行反插值：

$$P_n(x)=f(x_0)+(x-x_0)f[x_0,x_1]+(x-x_0)(x-x_1)f[x_0,x_1,x_2]+\cdots+(x-x_0)(x-x_1)\cdots(x-x_{n-1})f[x_0,x_1,\cdots,x_n] \tag{5.99}$$

令 $P_n(x)=c$ 后，将式(5.99)分解如下：

$$x=x_0+\frac{c-f(x_0)}{f[x_0,x_1]}-\frac{f[x_0,x_1,x_2]}{f[x_0,x_1]}(x-x_0)(x-x_1)-\cdots-$$

$$
\begin{aligned}
&\frac{f[x_0,x_1,\cdots,x_n]}{f[x_0,x_1]}(x-x_0)(x-x_1)\cdots(x-x_{n-1})\\
&=m_1+m_2(x-x_0)(x-x_1)+m_3(x-x_0)(x-x_1)(x-x_2)+\cdots+\\
&\quad m_n(x-x_0)(x-x_1)\cdots(x-x_{n-1})\\
&=\Phi(x)
\end{aligned}
\tag{5.100}
$$

其中 $m_1=x_0+\dfrac{c-f(x_0)}{f[x_0,x_1]}$， $m_2=-\dfrac{f[x_0,x_1,x_2]}{f[x_0,x_1]}$， $m_3=-\dfrac{f[x_0,x_1,x_2,x_3]}{f[x_0,x_1]}$，…，

$$m_n=-\frac{f[x_0,x_1,\cdots,x_n]}{f[x_0,x_1]}$$

对式(5.100)采用逐次逼近法进行迭代计算。

取 $x^{(0)}=m_1$ 后逐次迭代如下

$$
\begin{cases}
x^{(1)}=m_1+m_2(x^{(0)}-x_0)(x^{(0)}-x_1)\\
x^{(2)}=m_1+m_2(x^{(1)}-x_0)(x^{(1)}-x_1)+m_3(x^{(1)}-x_0)(x^{(1)}-x_1)(x^{(1)}-x_2)\\
\cdots\\
x^{(n-1)}=\Phi(x^{(n-2)})\\
x^{(n)}=\Phi(x^{(n-1)})\\
\cdots
\end{cases}
\tag{5.101}
$$

直到满足精度要求为止。

式(5.100)亦可按如下方式分解为

$$
\begin{aligned}
x=&x_0+\frac{c-f(x_0)}{f[x_0,x_1]+(x-x_1)f[x_0,x_1,x_2]+\cdots+(x-x_1)\cdots(x-x_{n-1})f[x_0,x_1,\cdots x_n]}\\
&=\Phi(x)
\end{aligned}
\tag{5.102}
$$

用逐次逼近法迭代如下

$$
\begin{cases}
x^{(0)}=x_0+\dfrac{c-f(x_0)}{f[x_0,x_1]}\\
x^{(1)}=x_0+\dfrac{c-f(x_0)}{f[x_0,x_1]+(x^{(0)}-x_1)f[x_0,x_1,x_2]}\\
x^{(2)}=x_0+\dfrac{c-f(x_0)}{f[x_0,x_1]+(x^{(1)}-x_1)f[x_0,x_1,x_2]+(x^{(1)}-x_1)(x^{(1)}-x_2)f[x_0,x_1,x_2,x_3]}\\
\cdots\\
x^{(n-1)}=\Phi(x^{(n-2)})\\
x^{(n)}=\Phi(x^{(n-1)})\\
\cdots
\end{cases}
\tag{5.103}
$$

直至达到精度要求为止。

对于等距节点的情形，可改用等距节点下的插值公式。以牛顿前插公式为例，设 $x_i=x_0+ih(i=0,1,2,\cdots,n)$且 $x=x_0+th(0<t<1)$，则与式(5.100)相应的公式为

$$t=\frac{c-y_0}{\Delta y_0}-\frac{\Delta^2 y_0}{2\Delta y_0}t(t-1)-\frac{\Delta^3 y_0}{3!\Delta y_0}t(t-1)(t-2)-\cdots-$$

$$\frac{\Delta^n y_0}{n!\Delta y_0}t(t-1)\cdots(t-n+1)$$
$$=\Phi(t) \tag{5.104}$$

与式(5.102)相应的公式为

$$t=\frac{c-y_0}{\Delta y_0+\frac{t-1}{2!}\Delta^2 y_0+\frac{(t-1)(t-2)}{3!}\Delta^3 y_0+\cdots+\frac{(t-1)(t-2)\cdots(t-n+1)}{n!}\Delta^n y_0}$$
$$=\Phi(t) \tag{5.105}$$

运用式(5.104)或式(5.105)建立迭代公式进行迭代，设达到精度要求时的近似值为 $\bar{t}$，则 $x\approx x_0+\bar{t}h$。

对于 $y=c$ 值，其对应的精确值设为 $x^*$，即

$$f(x^*)=c \tag{5.106}$$

又设满足 $P_n(x)=c$ 的 $x^*$ 的近似值为 $x$，则

$$f(x)=P_n(x)+R_n(x)=c+R_n(x) \tag{5.107}$$

两式相减得

$$f(x)-f(x^*)=R_n(x)$$
$$f'(\xi)(x-x^*)=R_n(x),\qquad \xi\in I$$

因此可得上述反插值法的余式估计公式

$$|R_n(y)|=|x-x^*|\leqslant\frac{|R_n(x)|}{m_1}\leqslant\frac{M_{n+1}}{m_1(n+1)!}\left|\prod\nolimits_{n+1}(x)\right| \tag{5.108}$$

其中
$$m_1\leqslant|f'(x)|,\qquad x\in I$$

**例 5.11**　用公式(5.104)解例 5.10。

**解**　建立差分表 5.15，按式(5.104)建立如下分解式：

$$t=\frac{0.98-0.982\ 154\ 32}{-0.003\ 957\ 71}-\frac{(-0.000\ 391\ 27)}{2\times(-0.003\ 957\ 71)}t(t-1)$$
$$=0.544\ 334\ 98-0.049\ 431\ 36t(t-1)$$
$$t_0=0.544\ 334\ 98$$
$$t_1=0.544\ 334\ 98-0.049\ 431\ 36\times0.544\ 334\ 98\times(0.544\ 334\ 98-1)$$
$$=0.556\ 595\ 66$$
$$t_2=0.556\ 534\ 49,\quad t_3=0.556\ 534\ 83,\quad t_4=0.556\ 534\ 83$$

最后得

$$x=1.76+0.556\ 534\ 83\times0.02=1.771\ 130\ 70$$

**表 5.15**

| $x$ | $y$ | $\Delta y$ | $\Delta^2 y$ |
|---|---|---|---|
| 1.76 | 0.982 154 32 | | |
| | | −0.003 957 71 | |
| 1.78 | 0.978 196 61 | | −0.000 391 27 |
| | | −0.004 348 98 | |
| 1.80 | 0.973 847 63 | | |

**例 5.12**　求方程 $x^5-5x+3=0$ 在[0,1]上的根。

**解**　取 $h=0.1$ 建立 $y=x^5-5x+3$ 的差分表 5.16。由表中可见，$0.6<\alpha<0.7$，取 $x_0=$

0.6，$c=0$，按式(5.104)建立如下的分解式：

$$t=\frac{0-0.077\ 76}{-0.409\ 69}+\frac{0.069\ 30}{2\times 0.409\ 69}t(t-1)+\frac{0.033\ 90}{6\times 0.409\ 69}t(t-1)(t-2)$$

或写成 $$t=0.189\ 80+0.084\ 58t(t-1)+0.013\ 79t(t-1)(t-2)$$

取 $t_0=0.189\ 80$，逐次迭代计算如下：

$$t_1=0.189\ 80+0.084\ 58\times 0.189\ 80\times(0.189\ 80-1)=0.176\ 79$$

$$t_2=0.189\ 80+0.084\ 58\times 0.176\ 79\times(0.176\ 79-1)+0.013\ 79\times 0.176\ 79\times(0.176\ 79-1)\times(0.176\ 79-2)=\Phi(t_1)=0.181\ 15$$

$$t_3=\Phi(t_2)=0.180\ 97$$

$$t_4=\Phi(t_3)=0.180\ 98$$

$$t_5=\Phi(t_4)=0.180\ 98$$

最后得 $$\alpha\approx 0.6+0.180\ 98\times 0.1=0.618\ 098$$

表 5.16

| $x$ | $y$ | $\Delta y$ | $\Delta^2 y$ | $\Delta^3 y$ | $\Delta^4 y$ |
|---|---|---|---|---|---|
| 0.4 | 1.010 24 | | | | |
| | | −0.478 99 | | | |
| 0.5 | 0.531 25 | | 0.025 50 | | |
| | | −0.453 49 | | 0.018 30 | |
| 0.6 | 0.077 76 | | 0.043 80 | | 0.007 20 |
| | | −0.409 69 | | 0.025 50 | |
| 0.7 | −0.331 93 | | 0.069 30 | | 0.008 40 |
| | | −0.340 39 | | 0.033 90 | |
| 0.8 | −0.672 32 | | 0.103 20 | | 0.009 60 |
| | | −0.237 19 | | 0.043 50 | |
| 0.9 | −0.909 51 | | 0.146 70 | | |
| | | −0.094 9 | | | |
| 1.0 | −1.000 00 | | | | |

### 6.2.2　求方程根的反插值法

(1) 单点弦截法

从插值的观点看，第二章叙述过的单点弦截法就是经过曲线 $f(x)$ 上一固定点 $(x_0,y_0)$ 与一活动点 $(x_n,y_n)$ 作一线性插值函数，求取 $y=0$ 时对应的 $x_{n+1}$ 值作为 $\alpha$ 近似值，插值函数为

$$P_1(x)=y_0+\frac{y_n-y_0}{x_n-x_0}(x-x_0)\tag{5.109}$$

令 $P_1(x_{n+1})=0$，代入上式解得单点弦截法的迭代公式

$$x_{n+1}=\frac{x_0f(x_n)-x_nf(x_0)}{f(x_n)-f(x_0)}\tag{5.110}$$

近似值 $x_{n+1}$ 对于根 $\alpha$ 的误差可按插值公式的余式来估计

$$f(\alpha)-P_1(\alpha)=\frac{f''(\xi)}{2!}(\alpha-x_0)(\alpha-x_n),\qquad \xi\in(x_0,x_n,\alpha)\tag{5.111}$$

因 $f(\alpha)=0$，$P_1(x_{n+1})=0$，所以上式可化为

$$\begin{aligned}P_1(x_{n+1})-P_1(\alpha)&=P'_1(\eta)(x_{n+1}-\alpha)\\&=\frac{f''(\xi)}{2}(x_0-\alpha)(x_n-\alpha),\qquad \eta\in(x_{n+1},\alpha)\end{aligned}\tag{5.112}$$

由此推知

$$\frac{x_{n+1}-\alpha}{x_n-\alpha}=\frac{f''(\xi)(x_0-\alpha)}{2P'_1(\eta)}\tag{5.113}$$

在收敛的情况下，有以下关系式成立

$$\lim_{n\to\infty}\left|\frac{x_{n+1}-\alpha}{x_n-\alpha}\right|=\left|\frac{f''(\alpha)}{2P_1'(\alpha)}(x_0-\alpha)\right|=c(\text{常数})\tag{5.114}$$

证得单点弦截法是线性收敛的。

(2) 双点弦截法

将式(5.109)～式(5.113)中的 $x_0$ 以 $x_{n-1}$ 代替后便得双点弦截法的有关公式。其中与式(5.113)相对应的公式为

$$\frac{x_{n+1}-\alpha}{(x_n-\alpha)(x_{n-1}-\alpha)}=\frac{f''(\xi)}{2P_1'(\eta)},\qquad \xi\in(x_{n-1},x_n,\alpha),\eta\in(x_{n+1},\alpha)\tag{5.115}$$

则有以下极限式

$$\lim_{n\to\infty}\frac{|x_{n+1}-\alpha|}{|x_n-\alpha||x_{n-1}-\alpha|}=\left|\frac{f''(\alpha)}{2P_1'(\alpha)}\right|=\left|\frac{f''(\alpha)}{2f'(\alpha)}\right|\tag{5.116}$$

设双点弦截法收敛的阶为 $p$，就有

$$\lim_{n\to\infty}\frac{|x_{n+1}-\alpha|}{|x_n-\alpha|^p}=c\tag{5.117}$$

及

$$\lim_{n\to\infty}\frac{|x_n-\alpha|}{|x_{n-1}-\alpha|^p}=c\tag{5.118}$$

成立。由式(5.118)得

$$\lim_{n\to\infty}|x_n-\alpha|=\lim_{n\to\infty}c|x_{n-1}-\alpha|^p\tag{5.119}$$

分别代入式(5.116)和式(5.117)后得

$$\lim_{n\to\infty}\frac{|x_{n+1}-\alpha|}{|x_{n-1}-\alpha|^{p+1}}=c\left|\frac{f''(\alpha)}{2f'(\alpha)}\right|\tag{5.120}$$

$$\lim_{n\to\infty}\frac{|x_{n+1}-\alpha|}{|x_{n-1}-\alpha|^{p^2}}=c^{p+1}\tag{5.121}$$

比较式(5.120)与式(5.121)得以下关系式

$$p^2=p+1,\quad \text{即 } p^2-p-1=0\tag{5.122}$$

及

$$c^{p+1}=c\left|\frac{f''(\alpha)}{2f'(\alpha)}\right|,\quad \text{即 } c=\left|\frac{f''(\alpha)}{2f'(\alpha)}\right|^{\frac{1}{p}}\tag{5.123}$$

由 $p^2-p-1=0$ 解得 $t_1=\frac{1+\sqrt{5}}{2}\approx1.618$， $t_2=\frac{1-\sqrt{5}}{2}\approx-0.618$。应取 $t_1>0$ 作为 $p$，从而证得双点弦截法的收敛阶为 1.618。

由式(5.115)可知

$$x_{n+1}-\alpha=\frac{f''(\xi)}{2p_1'(\eta)}(x_n-\alpha)(x_{n-1}-\alpha)\tag{5.124}$$

的符号与前两次近似值的误差符号及 $p'_1(\eta)$、$f''(\xi)$ 的符号变化有关，它们不规则的变化可能导致 $x_{n+1}-\alpha$ 符号的随机摆动。

(3) 密勒法（三点抛物线反插值求根法）

设 $\alpha\in[x_n,x_{n-2}]$，$f(x_n)f(x_{n-2})<0$，在$[x_n,x_{n-2}]$内取某点设为 $x_{n-1}$，在 $x_n$、$x_{n-1}$、$x_{n-2}$ 节点上使用牛顿基本差商公式得

$$P_2(x)=f(x_n)+(x-x_n)f[x_n,x_{n-1}]+(x-x_n)(x-x_{n-1})f[x_n,x_{n-1},x_{n-2}]\tag{5.125}$$

令 $P_2(x_{n+1})=0$ 代入上式得

$$f(x_n)+(x_{n+1}-x_n)f[x_n,x_{n-1}]+(x_{n+1}-x_n)(x_{n+1}-x_{n-1})f[x_n,x_{n-1},x_{n-2}]=0$$

$$f[x_n,x_{n-1},x_{n-2}](x_{n+1}-x_n)^2+\{f[x_n,x_{n-1}]+(x_n-x_{n-1})f[x_n,x_{n-1},x_{n-2}]\}\times (x_{n+1}-x_n)+f(x_n)=0$$

或改写为

$$a_n(x_{n+1}-x_n)^2+b_n(x_{n+1}-x_n)+c_n=0 \tag{5.126}$$

其中 $a_n=f[x_n,x_{n-1},x_{n-2}]$

$$b_n=f[x_n,x_{n-1}]+(x_n-x_{n-1})f[x_n,x_{n-1},x_{n-2}]=f[x_n,x_{n-1}]+a_n(x_n-x_{n-1})$$

$$c_n=f(x_n)$$

解得式(5.126)的两个根为

$$x_{n+1}=x_n+\frac{-b_n\pm\sqrt{b_n^2-4a_nc_n}}{2a_n}=x_n-\frac{2c_n}{b_n\pm\sqrt{b_n^2-4a_nc_n}} \tag{5.127}$$

两根中应选与 $x_n$ 接近的根为 $x_{n+1}$，即应取

$$x_{n+1}=x_n-\frac{2c_n}{b_n+\operatorname{sgn}(b_n)\sqrt{b_n^2-4a_nc_n}} \tag{5.128}$$

或

$$x_{n+1}=x_n-\frac{2c_n\operatorname{sgn}(b_n)}{|b_n|+\sqrt{b_n^2-4a_nc_n}} \tag{2.129}$$

以上求根方法称为密勒(Müller)法。其计算步骤如下。

① 给定精度 $\varepsilon>0$，初始值 $x_2$、$x_1$、$x_0$，计算 $f(x_2)$、$f(x_1)$、$f(x_0)$。

② 计算

$$a_2=f[x_2,x_1,x_0]$$

$$b_2=f[x_2,x_1]+a_2(x_2-x_1)$$

$$c_2=f(x_2)$$

$$x_3=x_2-\frac{2c_2\operatorname{sgn}(b_2)}{|b_2|+\sqrt{b_2^2-4a_2c_2}}$$

③ 若 $|x_3-x_2|<\varepsilon$，则 $\alpha=x_3$；否则，下标增 1 后转①。

用与双点弦截法中类似方法可导得密勒法误差估计的极限式

$$\lim_{n\to\infty}\left|\frac{x_{n+1}-\alpha}{(x_n-\alpha)(x_{n-1}-\alpha)(x_{n-2}-\alpha)}\right|=\left|-\frac{1}{6}\frac{f'''(\alpha)}{f'(\alpha)}\right|=\left|\frac{f'''(\alpha)}{6f'(\alpha)}\right| \tag{5.130}$$

设密勒法收敛的阶为 $p$，就有

$$\lim_{n\to\infty}\frac{|x_{n+1}-\alpha|}{|x_n-\alpha|^p}=c \tag{5.131}$$

$$\lim_{n\to\infty}\frac{|x_n-\alpha|}{|x_{n-1}-\alpha|^p}=c \tag{5.132}$$

$$\lim_{n\to\infty}\frac{|x_{n-1}-\alpha|}{|x_{n-2}-\alpha|^p}=c \tag{5.133}$$

成立。由式(5.133)解得

$$\lim_{n\to\infty}|x_{n-1}-\alpha|=\lim_{n\to\infty}c|x_{n-2}-\alpha|^p \tag{5.134}$$

代入式(5.132)得

$$\lim_{n\to\infty}|x_n-\alpha|=\lim_{n\to\infty}c|x_{n-1}-\alpha|^p=\lim_{n\to\infty}c^{p+1}|x_{n-2}-\alpha|^{p^2} \tag{5.135}$$

将上式代入式(5.131)得

$$\lim_{n\to\infty}\frac{|x_{n+1}-\alpha|}{|x_{n-2}-\alpha|^{p^3}}=c^{p^2+p+1} \tag{5.136}$$

再将式(5.135)、式(5.134)代入式(5.130)得

$$\lim_{n\to\infty}\frac{|x_{n+1}-\alpha|}{|x_{n-2}-\alpha|^{p^2+p+1}}=c^{p+2}\left|\frac{f'''(\alpha)}{6f'(\alpha)}\right| \tag{5.137}$$

比较式(5.136)与式(5.137)得

$$p^3=p^2+p+1,\ c^{p^2+p+1}=c^{p+2}\left|\frac{f'''(\alpha)}{6f'(\alpha)}\right|$$

获得解

$$p=1.839\approx 1.84$$

$$c=\left|\frac{f'''(\alpha)}{6f'(\alpha)}\right|^{\frac{1}{p^2-1}} \tag{5.138}$$

(4) 继续采用节点 $x_n, x_{n-1}, \cdots, x_{n-i}$ 上的插值函数 $P_i(x)(i\geqslant 3)$ 近似 $f(x)$，令

$$P_i(x_{n+1})=0 \tag{5.139}$$

因 $i\geqslant 3$，从式(5.139)中直接解出 $x_{n+1}$ 的方法已不可取，只能用迭代法求解方程(5.139)得 $x_{n+1}$。然后再以 $x_{n+1}, x_n, \cdots, x_{n-i+1}$ 为节点建立 $P_i(x)$，令

$$P_i(x_{n+2})=0 \tag{5.140}$$

仿上求出 $x_{n+2}$，如此重复直到 $|x_{n+1}-x_n|<\varepsilon$ 满足为止，则 $\alpha=x_{n+1}$。

用类似方法可推算 $i=3,4,\cdots$ 的收敛阶数。下面列出 $i=1,2,3,4$ 时的 $p$ 值，见表 5.17。

**表 5.17**

| $P_i(x)$ | $P_1(x)$ | | $P_2(x)$ | $P_3(x)$ | $P_4(x)$ |
|---|---|---|---|---|---|
| | 单点弦截法 | 双点弦截法 | | | |
| $p$ | 1.000 | 1.618 | 1.839 | 1.928 | 1.966 |

**例 5.13**　用密勒法求方程

$$f(x)=xe^x-1=0$$

在 $x_0=0.5$ 附近的根。

**解**：取 $x_2=0.6, x_1=0.55, x_0=0.5$，计算相应的函数值与差商见表 5.18，则有

$$a_2=2.2108$$

$$b_2=2.79964+2.2108\times 0.05=2.91018$$

$$c_2=0.093271$$

$$x_3=0.6-\frac{2\times 0.093271}{2.91018+\sqrt{2.91018^2-4\times 2.2108\times 0.093271}}=0.56713$$

$$f(x_3)=-0.000037$$

$$f[x_3,x_2]=2.838698$$

$$f[x_3,x_2,x_1]=2.280093$$

则有

$$a_3=2.280093$$

$$b_3=2.838698+2.280093\times(0.56713-0.6)=2.763751$$

$$c_3=-0.000037$$

$$x_4 = 0.567\,13 - \frac{2\times(-0.000\,037)}{2.763\,751 + \sqrt{2.763\,751^2 - 4\times 2.280\,093\times(-0.000\,037)}}$$
$$= 0.567\,14$$
$$f(x_4) = -0.000\,009$$
$$f[x_4, x_3] = 2.8$$
$$f[x_4, x_3, x_2] = 1.177\,663$$

则有
$$a_4 = 1.177\,663$$
$$b_4 = 2.8 + 1.177\,663\times 0.000\,01 = 2.800\,012$$
$$c_4 = -0.000\,009$$
$$x_5 = 0.567\,14 - \frac{2\times(-0.000\,009)}{2.800\,012 + \sqrt{2.800\,012^2 - 4\times 1.177\,663\times(-0.000\,009)}}$$
$$= 0.567\,14$$

最后把上述计算结果列于表 5.18 中。

表 5.18

| $n$ | $x_n$ | $f(x_n)$ | $f[x_n, x_{n+1}]$ | $f[x_n, x_{n+1}, x_{n+2}]$ |
|---|---|---|---|---|
| 0 | 0.5 | −0.175 639 | 2.578 560 | 2.210 8 |
| 1 | 0.55 | −0.046 711 | 2.799 64 | 2.280 093 |
| 2 | 0.60 | 0.093 271 | 2.838 698 | 1.177 663 |
| 3 | 0.567 13 | −0.000 037 | 2.8 | |
| 4 | 0.567 14 | −0.000 009 | | |
| 5 | 0.567 14 | | | |

# §7 埃尔米特插值多项式

以上介绍的插值公式只要求插值多项式在插值节点上取给定的函数值。有时还要求插值多项式在某些节点或全部节点上具有给定的导数值。设在节点 $x_0, x_1, \cdots, x_n$ 上已知下列函数值与导数值：

$$\begin{cases} y_0, y'_0, y''_0, \cdots, y_0^{(m_0)} \\ y_1, y'_1, y''_1, \cdots, y_1^{(m_1)} \\ \cdots \\ y_n, y'_n, y''_n, \cdots, y_n^{(m_n)} \end{cases} \tag{5.141}$$

今要求这样的多项式 $P_m(x)$，使 $P_m(x)$ 在节点 $x_0, x_1, \cdots, x_n$ 上取(5.141)中的给定值。式(5.141)中的条件总数为

$$\sum_{i=0}^{n}(m_i + 1) = m + 1 \tag{5.142}$$

利用以上条件可以确定一个 $m$ 次多项式

$$P_m(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m \tag{5.143}$$

使满足式(5.141)，称多项式 $P_m(x)$ 为埃尔米特插值多项式。它应满足下式

$$\begin{cases}\sum_{k=0}^{m} x_i^k a_k = y_i \\ \sum_{k=1}^{m} k x_i^{k-1} a_k = y_i' \qquad (i=0,1,2,\cdots,n) \\ \cdots \\ \sum_{k=m_i}^{m} k(k-1)\cdots(k-\overline{m_i-1}) x_i^{k-m_i} a_k = y_i^{(m_i)}\end{cases} \tag{5.144}$$

式(5.144)是一个以 $a_0, a_1, \cdots, a_m$ 为未知量的线性方程组。因函数组$\{1,x,x^2,\cdots,x^m\}$在点集 $x_0, x_1,\cdots,x_n$ 上线性无关,故式(5.144)的系数行列式不等于零,从而知 $a_0,a_1,\cdots,a_m$ 唯一确定。为了免除求解式(5.144)的困难,有了上面唯一性的证明,就可研制其他简捷的算法求取 $P_m(x)$。

## 7.1　牛顿型埃尔米特插值公式

联想到一个节点上的各阶导数值与重节点上的差商值有关系式(5.22),即

$$\begin{cases} f[x_i,x_i] = \dfrac{f'(x_i)}{1!} = y'_i \\ f[x_i,x_i,x_i] = \dfrac{f''(x_i)}{2!} = y''_i/2! \\ \cdots \\ f[\underbrace{x_i,x_i,\cdots,x_i}_{m_i+1\text{个}}] = \dfrac{f^{(m_i)}(x_i)}{m_i!} = y_i^{(m_i)}/m_i! \end{cases} \tag{5.145}$$

因此若在差商表中,对每一节点 $x_i$ 按其导数的最大阶数 $m_i$ 将它的个数分别增加到 $m_i+1$ 个后,再利用式(5.145)的关系,就可将(5.141)的全部条件加入到差商表中,而 $P_m(x)$的求取问题就可归结为在下述 $m+1$ 个互异节点组

$$\underbrace{x_0,x_0,\cdots,x_0}_{m_0+1\text{个}},\underbrace{x_1,x_1,\cdots,x_1}_{m_1+1\text{个}},\cdots,\underbrace{x_n,x_n,\cdots,x_n}_{m_n+1\text{个}} \tag{5.146}$$

上的插值问题。按牛顿基本差商公式可将它表为

$$\begin{aligned} P_m(x) = {} & f(x_0) + (x-x_0)f[x_0,x_0] + (x-x_0)^2 f[x_0,x_0,x_0] + \cdots + \\ & (x-x_0)^{m_0+1} f[\underbrace{x_0,\cdots,x_0}_{m_0+1},x_1] + \\ & (x-x_0)^{m_0+1}(x-x_1) f[\underbrace{x_0,\cdots,x_0}_{m_0+1},x_1,x_1] + \cdots + \\ & (x-x_0)^{m_0+1}(x-x_1)^{m_1+1} f[\underbrace{x_0,\cdots x_0}_{m_0+1},\underbrace{x_1,\cdots,x_1}_{m_1+1},x_2] + \cdots + \\ & (x-x_0)^{m_0+1}(x-x_1)^{m_1+1}\cdots(x-x_{n-1})^{m_{n-1}+1}(x-x_n)^{m_n} \times \\ & f[\underbrace{x_0,\cdots,x_0}_{m_0+1},\cdots,\underbrace{x_{n-1},\cdots,x_{n-1}}_{m_{n-1}+1},\underbrace{x_n,\cdots,x_n}_{m_n+1}] \end{aligned} \tag{5.147}$$

称公式(5.147)为重节点插值多项式,其余式为

$$\begin{aligned} R_m(x) = f(x) - P_m(x) &= f[\underbrace{x_0,\cdots,x_0}_{m_0+1},\cdots,\underbrace{x_n,\cdots,x_n}_{m_n+1},x]\prod_{i=0}^{n}(x-x_i)^{m_i+1} \\ &= \frac{f^{(m+1)}(\xi)}{(m+1)!}\prod_{i=0}^{n}(x-x_i)^{m_i+1} \end{aligned} \tag{5.148}$$

其中 $\xi$ 属于由 $x_0, x_1, \cdots, x_n, x$ 所界的区间内。

**例 5.14** 已知数值表表 5.19,求符合表值的埃尔米特插值多项式。

**表 5.19**

| $x$ | $y$ | $y'$ | $y''$ |
|---|---|---|---|
| 0 | 3 | 4 | |
| 1 | 5 | 6 | 7 |

**解** 因 $x=0$ 时,$y$ 的导数最高阶数为 1,在差商表中应重复取作二个节点;而 $x=1$ 时,$y$ 的导数最高阶为 2,则在差商表中应重复取作三个节点。经这样处理后,按以下节点 0,0,1,1,1 建立差商表 5.20。

**表 5.20**

| $x$ | $y$ | 一阶差商 | 二阶差商 | 三阶差商 | 四阶差商 |
|---|---|---|---|---|---|
| 0 | 3 | | | | |
| | | $f[0,0]=f'(0)=4$ | | | |
| 0 | 3 | | $f[0,0,1]=\frac{2-4}{1-0}=-2$ | | |
| | | $f[0,1]=\frac{5-3}{1-0}=2$ | | $f[0,0,1,1]=\frac{4+2}{1-0}=6$ | |
| 1 | 5 | | $f[0,1,1]=\frac{6-2}{1-0}=4$ | | $f[0,0,1,1,1]=\frac{-0.5-6}{1-0}=-6.5$ |
| | | $f[1,1]=f'(1)=6$ | | $f[0,1,1,1]=\frac{3.5-4}{1-0}=-0.5$ | |
| 1 | 5 | | $f[1,1,1]=\frac{f''(1)}{2}=3.5$ | | |
| | | $f[1,1]=f'(1)=6$ | | | |
| 1 | 5 | | | | |

按牛顿基本差商公式得

$$
\begin{aligned}
P_4(x) &= 3+(x-0)\times 4+(x-0)^2\times(-2)+ \\
&\quad (x-0)^2(x-1)\times 6+(x-0)^2(x-1)^2\times(-6.5) \\
&= 3+4x-2x^2+6x^2(x-1)-6.5x^2(x-1)^2 \\
&= -6.5x^4+19x^3-14.5x^2+4x+3
\end{aligned}
$$

## 7.2 降阶型埃尔米特插值公式

这是一种把高次插值多项式逐次转化为低次插值多项式的求取方法,以下面的实例具体说明之。

**例 5.15** 已知数值表表 5.21,求符合表值的埃尔米特插值多项式。

**表 5.21**

| $x$ | $y$ | $y'$ | $y''$ |
|---|---|---|---|
| 0 | 2 | 2 | −10 |
| 1 | 1 | −1 | 0 |
| 2 | 2 | 6 | 30 |

**解** 根据表中条件可以确定一个 $m=9-1=8$ 次的埃尔米特插值多项式 $P_8(x)$。首先利用表 5.22。

表 5.22

| $x$ | 0 | 1 | 2 |
|---|---|---|---|
| $y$ | 2 | 1 | 2 |

建立插值公式得

$$L_{21}(x) = x^2 - 2x + 2$$

则 $P_8(x)$ 可表为

$$P_8(x) = L_{21}(x) + (x-0)(x-1)(x-2)P_5(x) \tag{5.149}$$

对 $P_8(x)$ 求导两次得

$$P'_8(x) = (2x-2) + (3x^2 - 6x + 2)P_5(x) + (x-0)(x-1)(x-2)P_5'(x) \tag{5.150}$$

$$P''_8(x) = 2 + (6x-6)P_5(x) + 2(3x^2 - 6x + 2)P_5'(x) + (x-0)(x-1)(x-2)P''_5(x) \tag{5.151}$$

利用表值条件 $P'_8(0)=2, P'_8(1)=-1, P'_8(2)=6$ 可分别解得

$$P_5(0) = 2,\quad P_5(1) = 1,\quad P_5(2) = 2$$

继续利用条件 $P''_8(0)=-10, P''_8(1)=0, P''_8(2)=30$ 可分别解得

$$P'_5(0) = 0,\quad P'_5(1) = 1,\quad P'_5(2) = 4$$

则 $P_5(x)$ 应满足表 5.23。

表 5.23

| $x$ | 0 | 1 | 2 |
|---|---|---|---|
| $P_5(x)$ | 2 | 1 | 2 |
| $P'_5(x)$ | 0 | 1 | 4 |

再利用表 5.24 建立插值公式得

$$L_{22}(x)=x^2-2x+2$$

表 5.24

| $x$ | 0 | 1 | 2 |
|---|---|---|---|
| $P_5(x)$ | 2 | 1 | 2 |

则 $P_5(x)$ 可表为

$$P_5(x) = L_{22}(x) + (x-0)(x-1)(x-2)P_2(x) \tag{5.152}$$

对 $P_5(x)$ 求导得

$$P'_5(x) = (2x-2) + (3x^2 - 6x + 2)P_2(x) + (x-0)(x-1)(x-2)P'_2(x)$$

令 $P'_5(0)=0, P'_5(1)=1, P'_5(2)=4$ 可分别解得

$$P_2(0) = 1,\quad P_2(1) = -1,\quad P_2(2) = 1$$

建立表 5.25。

表 5.25

| $x$ | 0 | 1 | 2 |
|---|---|---|---|
| $P_2(x)$ | 1 | $-1$ | 1 |

按表 5.25 求得插值公式

$$P_2(x) = L_{23}(x) = 2x^2 - 4x + 1 \tag{5.153}$$

将式(5.153)代入式(5.152)得

$$\begin{aligned} P_5(x) &= (x^2 - 2x + 2) + (x^3 - 3x^2 + 2x)(2x^2 - 4x + 1) \\ &= 2x^5 - 10x^4 + 17x^3 - 10x^2 + 2 \end{aligned}$$

再代入式(5.149)得

$$\begin{aligned} P_8(x) &= (x^2 - 2x + 2) + (x^3 - 3x^2 + 2x)(2x^5 - 10x^4 + 17x^3 - 10x^2 + 2) \\ &= 2x^8 - 16x^7 + 51x^6 - 81x^5 + 64x^4 - 18x^3 - 5x^2 + 2x + 2 \end{aligned}$$

## 7.3 拉格朗日型埃尔米特插值公式

最常见的一种埃尔米特插值问题是要求插值多项式与 $f(x)$ 在节点上具有相同的函数值与一阶导数值。以下我们来导出这种插值多项式。

设已知在插值节点 $x_i(i=0,1,2,\cdots,n)$ 上的函数值为 $y_i=f(x_i)$ 以及一阶导数值为 $y'_i=f'(x_i)$，要求构造满足上述条件的插值多项式。因条件数为 $2n+2$ 个，可以确定一个次数不超过 $2n+1$ 次的插值多项式 $P_{2n+1}(x)$，使满足

$$\begin{cases} P_{2n+1}(x_i) = y_i \\ P'_{2n+1}(x_i) = y'_i \end{cases} \quad (i = 0,1,2,\cdots,n) \tag{5.154}$$

仿照拉格朗日插值公式中插值基函数的功能，设

$$P_{2n+1}(x) = \sum_{i=0}^{n} [\alpha_i(x) y_i + \beta_i(x) y'_i] \tag{5.155}$$

要求 $\alpha_i(x)$ 与 $\beta_i(x)$ 分别满足

$$\begin{cases} \alpha_i(x_j) = \begin{cases} 0, & j \neq i \\ 1, & j = i \end{cases} \quad (i,j = 0,1,2,\cdots,n) \\ \alpha'_i(x_j) = 0 \end{cases} \tag{5.156}$$

与

$$\begin{cases} \beta'_i(x_j) = \begin{cases} 0, & j \neq i \\ 1, & j = i \end{cases} \\ \beta_i(x_j) = 0 \quad (i,j = 0,1,2,\cdots,n) \end{cases} \tag{5.157}$$

这样确定的 $\alpha_i(x)$ 与 $\beta_i(x)$ 称为埃尔米特插值基函数。显然这种基函数一旦确定下来，就能使式(5.155)中的 $P_{2n+1}(x)$ 满足条件式(5.154)。下面导出 $\alpha_i(x)$ 与 $\beta_i(x)$ 的具体公式。

先确定 $\beta_i(x)$，因为

$$\beta_i(x_j) = 0, \quad \beta'_i(x_j) = 0 \qquad (j \neq i) \tag{5.158}$$

所以 $\beta_i(x)$ 以 $x_j(j\neq i)$ 为二重零点，又因为

$$\beta_i(x_i) = 0, \quad \beta'_i(x_i) = 1 \tag{5.159}$$

所以 $\beta_i(x)$ 以 $x_i$ 为一重零点。由于 $\beta_i(x)$ 为次数不超过 $2n+1$ 次的多项式，又按拉格朗日插值

公式的基函数的功能,可令

$$\beta_i(x) = c_i(x-x_i)l_i^2(x) \qquad (c_i \text{ 为常数}) \tag{5.160}$$

其中

$$l_i(x) = \frac{(x-x_0)(x-x_1)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)}{(x_i-x_0)(x_i-x_1)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_n)} \tag{5.161}$$

为 $n$ 次多项式且满足 $l_i(x_j)=0(j\neq i)$ 及 $l_i(x_i)=1$。由式(5.160)得

$$\beta'_i(x) = c_i[l_i^2(x)+2(x-x_i)l_i(x)l'_i(x)]$$

代入 $l_i(x_i)=1$ 及 $\beta'_i(x_i)=1$ 解得 $c_i=1$。所以

$$\beta_i(x) = (x-x_i)l_i^2(x) \tag{5.162}$$

为确定 $\alpha_i(x)$,仿上分析知 $\alpha_i(x_j)=\alpha'(x_j)=0(j\neq i)$,所以 $x_j(j\neq i)$ 为 $\alpha_i(x)$ 的二重零点。又 $\alpha_i(x_i)=1$,所以 $x_i$ 不是 $\alpha_i(x)$ 的零点,由于 $\alpha_i(x)$ 为小于 $2n+1$ 次的多项式,故可令

$$\alpha_i(x) = (ax+b)l_i^2(x) \tag{5.163}$$

其中常数 $a,b$ 由条件 $\alpha_i(x_i)=1$ 及 $\alpha'_i(x_i)=0$ 确定

$$\begin{cases} ax_i+b=1 \\ a+2l'_i(x_i)=0 \end{cases}$$

解得

$$\begin{cases} a=-2l'_i(x_i) \\ b=1+2x_il'_i(x_i) \end{cases} \tag{5.164}$$

因

$$\ln l_i(x) = \ln \frac{(x-x_0)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)}{(x_i-x_0)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_n)}$$

$$= \sum_{\substack{j=0\\j\neq i}}^{n}\ln(x-x_j) - \sum_{\substack{j=0\\j\neq i}}^{n}\ln(x_i-x_j)$$

两边求导得

$$\frac{l'_i(x)}{l_i(x)} = \sum_{\substack{j=0\\j\neq i}}^{n}\frac{1}{x-x_j}$$

$$l'_i(x_i) = l_i(x_i)\cdot\sum_{\substack{j=0\\j\neq i}}^{n}\frac{1}{x_i-x_j} = \sum_{\substack{j=0\\j\neq i}}^{n}\frac{1}{x_i-x_j} \tag{5.165}$$

代入式(5.163)得

$$\begin{aligned} \alpha_i(x) &= [-2l_i'(x_i)x+1+2x_il_i'(x_i)]l_i^2(x) \\ &= [1-2(x-x_i)l'_i(x_i)]l_i^2(x) \\ &= [1-2(x-x_i)\sum_{\substack{j=0\\j\neq i}}^{n}\frac{1}{x_i-x_j}]l_i^2(x) \end{aligned} \tag{5.166}$$

这样式(5.155)可表为

$$P_{2n+1}(x) = \sum_{i=0}^{n}\left\{y_i-(x-x_i)[2y_i\sum_{\substack{j=0\\j\neq i}}^{n}\frac{1}{x_i-x_j}-y_i']\right\}l_i^2(x) \tag{5.167}$$

其余式为

$$R_{2n+1}(x) = \frac{f^{(2n+2)}(\xi)}{(2n+2)!}[(x-x_0)(x-x_1)\cdots(x-x_n)]^2 \tag{5.168}$$

其中 $\xi$ 属于 $x_0,x_1,\cdots,x_n,x$ 所界的区间内。公式(5.167)的几何意义是,曲线 $P_{2n+1(x)}$ 与 $f(x)$ 在插值节点处有公共切线。

特别地，当 $n=1$ 时，式(5.155)成为

$$
\begin{aligned}
P_3(x) &= \sum_{i=0}^{1}\alpha_i(x)y_i + \sum_{i=0}^{1}\beta_i(x)y'_i \\
&= \left(1-2\frac{x-x_0}{x_0-x_1}\right)l_0^2(x)\cdot y_0 + \left(1-2\frac{x-x_1}{x_1-x_0}\right)l_1^2(x)\cdot y_1 + \\
&\quad (x-x_0)l_0^2(x)\cdot y'_0 + (x-x_1)l_1^2(x)\cdot y'_1
\end{aligned} \tag{5.169}
$$

它满足以下插值条件

$$P_3(x_0)=y_0,\quad P_3(x_1)=y_1,\quad P'_3(x_0)=y'_0,\quad P'_3(x_1)=y'_1$$

式(5.169)常称为三次埃尔米特插值多项式。

其余式为

$$R_3(x)=\frac{f^{(4)}(\xi)}{4!}(x-x_0)^2(x-x_1)^2,\qquad \xi\in[x_0,x_1]$$

因 $$|(x-x_0)(x-x_1)|\leqslant|(x-x_0)(x-x_1)|_{x=\frac{x_0+x_1}{2}}=\frac{(x_1-x_0)^2}{4}$$

所以有以下三次埃尔米特插值公式的余式估计

$$|R_3(x)|\leqslant\frac{M_4}{384}(x_1-x_0)^4,\quad M_4=\max_{x\in[x_0,x_1]}|f^{(4)}(x)| \tag{5.170}$$

如果把区间$[a,b]$分成 $n$ 个子区间$[x_i,x_{i+1}]$，$i=0,1,2,\cdots,n-1$，在每个子区间上用上述三次埃尔米特插值多项式来近似函数 $y=f(x)$，则称这种插值函数为分段三次埃尔米特插值函数。

仿此，还可以得到 $m_i=2(i=0,1,2,\cdots,n)$时的拉格朗日型插值公式如下

$$P_{3n+2}(x)=\sum_{i=0}^{n}[\alpha_i(x)y_i+\beta_i(x)y'_i+\gamma_i(x)y''_i] \tag{5.171}$$

其中

$$
\begin{cases}
\alpha_i(x)=[1-3a_i(x-x_i)+b_i(x-x_i)^2]\cdot l_i^3(x) \\
\beta_i(x)=(x-x_i)[1-3a_i(x-x_i)]\cdot l_i^3(x) \\
\gamma_i(x)=\dfrac{1}{2}(x-x_i)^2\cdot l_i^3(x) \\
l_i(x)=\dfrac{(x-x_0)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)}{(x_i-x_0)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_n)} \\
a_i=l'_i(x_i)=\sum\limits_{\substack{j=0\\j\neq i}}^{n}\dfrac{1}{x_i-x_j} \\
b_i=6a_i^2-\dfrac{3}{2}c_i \\
c_i=l''_i(x_i)=a_i^2-\sum\limits_{\substack{j=0\\j\neq i}}^{n}\dfrac{1}{(x_i-x_j)^2} \qquad (i=0,1,2,\cdots,n)
\end{cases} \tag{5.172}
$$

余式为

$$R_{3n+2}(x)=f(x)-P_{3n+2}(x)=\frac{f^{(3n+3)}(\xi)}{(3n+3)!}[(x-x_0)(x-x_1)\cdots(x-x_n)]^3 \tag{5.173}$$

**例 5.16** 求符合表 5.26 的埃尔米特插值多项式。

表 5.26

| $x$ | 2.2 | 2.4 |
| --- | --- | --- |
| $y=\ln x$ | 0.788 46 | 0.875 47 |
| $y'=\frac{1}{x}$ | 0.454 55 | 0.416 67 |

计算 $f(2.3)$的近似值。

**解** 按式(5.169)建立以下多项式

$$P_3(x)=\left(1-2\times\frac{x-2.2}{2.2-2.4}\right)\left(\frac{x-2.4}{2.2-2.4}\right)^2\times 0.788\,46+\left(1-2\times\frac{x-2.4}{2.4-2.2}\right)\left(\frac{x-2.2}{2.4-2.2}\right)^2\times 0.875\,47+(x-2.2)\left(\frac{x-2.4}{2.2-2.4}\right)^2\times 0.454\,55+(x-2.4)\left(\frac{x-2.2}{2.4-2.2}\right)^2\times 0.416\,67$$

以 $x=2.3$ 代入上式计算得 $\ln 2.3\approx P_3(2.3)=0.832\,91$。

由于

$$(\ln x)^{(4)}=-\frac{6}{x^4}$$

所以其余式估计为

$$|R_3(2.3)|\leqslant\frac{1}{4!}\cdot\frac{6}{2.2^4}(2.3-2.2)^2(2.3-2.4)^2\approx 1.067\times 10^{-6}$$

## 7.4 特殊情形的埃尔米特插值

在埃尔米特插值条件中，可以允许函数值与导数值的个数不等，这时可以前述几种类型的埃尔米特插值多项式为基础，运用待定系数法求出满足插值条件的多项式，下面举例说明之。

**例 5.17** 求符合表 5.27 的埃尔米特插值多项式。

表 5.27

| $x$ | 0 | 1 | 4 |
| --- | --- | --- | --- |
| $y$ | 0 | | 4 |
| $y'$ | 2 | 1 | |

**解** 根据表值，$x=0,1$ 时 $y$ 的最高阶导数的阶数均为 1，在差商表中应分别重复取作两个节点，以下按 0,0,1,1,4 的节点顺序建立差商表 5.28。

因条件数为 4，可确定出 $P_3(x)$，则其三阶差商为常数，因得

$$-2y_1+3=\frac{2y_1-2}{9}$$

解得 $y_1=1.45$。利用差商表建立牛顿基本差商公式得

$$P_3(x)=0+2(x-0)+(x-0)(x-0)(y_1-2)+(x-0)(x-0)(x-1)(-2y_1+3)$$

以 $y_1=1.45$ 代入得

$$P_3(x)=0.1x^3-0.65x^2+2x$$

表 5.28

| $x$ | $y$ | 一阶差商 | 二阶差商 | 三阶商 |
| --- | --- | --- | --- | --- |
| 0 | 0 | | | |
| | | $\frac{y'_0}{1!}=2$ | | |
| 0 | 0 | | $\frac{y_1-2}{1-0}=y_1-2$ | |
| | | $\frac{y_1-0}{1-0}=y_1$ | | $\frac{(1-y_1)-(y_1-2)}{1-0}=-2y_1+3$ |
| 1 | $y_1$ | | $\frac{1-y_1}{1-0}=1-y_1$ | |
| | | $\frac{y'_1}{1!}=1$ | | $\left[\frac{1-y_1}{9}-(1-y_1)\right]/(4-0)=\frac{2y_1-2}{9}$ |
| 1 | $y_1$ | | $\left[\frac{4-y_1}{3}-1\right]/(4-1)=\frac{1-y_1}{9}$ | |
| | | $\frac{4-y_1}{4-1}=\frac{4-y_1}{3}$ | | |
| 4 | 4 | | | |

亦可按节点 0,4 建立差商表表 5.29。

表 5.29

| $x$ | $y$ | 一阶差商 |
| --- | --- | --- |
| 0 | 0 | |
| | | 1 |
| 4 | 4 | |

按牛顿基本差商公式得

$$P_1(x)=0+(x-0)\times 1=x$$

令

$$P_3(x)=P_1(x)+(x-0)(x-4)(ax+b)$$

$$P'_3(x)=1+(2x-4)(ax+b)+x(x-4)\cdot a$$

利用条件 $P'_3(0)=2$ 及 $P'_3(1)=1$ 解得

$$a=0.1,\quad b=-0.25$$

所以 $P_3(x)=x+x(x-4)(0.1x-0.25)=0.1x^3-0.65x^2+2x$

**例 5.18** 求符合表 5.30 的埃尔米特插值多项式。

表 5.30

| $x$ | 0 | 1 | 4 |
| --- | --- | --- | --- |
| $y$ | 0 | 2.5 | 4 |
| $y'$ | 2 | 1 | |

**解** ① 若按牛顿型埃尔米特插值多项式为基础,可据表值建立差商表 5.31。

表 5.31

| $x$ | $y$ | 一阶差商 | 二阶差商 |
| --- | --- | --- | --- |
| 0 | 0 | | |
| | | 2.5 | |
| 1 | 2.5 | | −0.5 |
| | | 0.5 | |
| 4 | 4 | | |

利用差商表建立牛顿基本差商公式

$$P_2(x)=0+2.5(x-0)-0.5(x-0)(x-1)=-0.5x^2+3x$$

令

$$P_4(x)=P_2(x)+(x-0)(x-1)(x-4)P_1(x) \tag{5.174}$$

$$P'_4(x)=(-x+3)+(3x^2-10x+4)P_1(x)+(x-0)(x-1)(x-4)P'_1(x)$$

由条件 $P'_4(0)=2$ 及 $P'_4(1)=1$ 解得

$$\begin{cases}P_1(0)=-0.25\\ P_1(1)=\dfrac{1}{3}\end{cases} \tag{5.175}$$

令 $P_1(x)=ax+b$，利用式(5.175)可解得

$$a=\frac{7}{12},\quad b=-0.25$$

则

$$P_1(x)=\frac{1}{12}(7x-3) \tag{5.176}$$

代入式(5.174)得

$$P_4(x)=(-0.5x^2+3x)+(x-0)(x-1)(x-4)\cdot\frac{1}{12}(7x-3)$$

$$=\frac{7}{12}x^4-\frac{38}{12}x^3+\frac{37}{12}x^2+2x \tag{5.177}$$

② 若按拉格朗日型埃尔米特插值多项式为基础，可建立以下插值公式

$$P_3(x)=\left(1-2\frac{x-0}{0-1}\right)\left(\frac{x-1}{0-1}\right)^2\times 0+\left(1-2\frac{x-1}{1-0}\right)\left(\frac{x-0}{1-0}\right)^2\times 2.5+$$

$$(x-0)\left(\frac{x-1}{0-1}\right)^2\times 2+(x-1)\left(\frac{x-0}{1-0}\right)^2\times 1$$

$$=-2x^3+2.5x^2+2x \tag{5.178}$$

令

$$P_4(x)=P_3(x)+c(x-0)^2(x-1)^2$$

利用条件 $P_4(4)=4$ 解得 $c=\frac{7}{12}$，因而

$$P_4(x)=(-2x^3+2.5x^2+2x)+\frac{7}{12}x^2(x-1)^2$$

$$=\frac{7}{12}x^4-\frac{38}{12}x^3+\frac{37}{12}x^2+2x$$

对于函数值个数与导数值个数不等的埃尔米特插值问题，可以类似于一般的埃尔米特插值问题得到其插值多项式的唯一性及其余式的表达式。

# §8 三次样条插值

## 8.1 样条的概念

采用分段线性插值与分段二次插值，可以构造一个整个连续的函数，而采用分段三次埃尔米特插值可构造一个整体上具有一阶连续导数的插值函数。能否把整体光滑度再加以提高以满足应用上的高精度要求呢？在原理上，应用高次埃尔米特插值多项式可以达到上述目的，但

它必须提供已知节点上的导数值为前提条件，这在实际上是较难满足的。能否在只给出节点对应的函数值的情况下构造出一个整体上具有二阶连续导数的插值函数呢？这就是本节要介绍的三次样条插值函数。

样条这一名词来源于工程中的样条曲线。绘图员为了将一些指定点(样点)联结成一条光滑曲线，经常用一条富有弹性的细长金属条(样条)把相近的几点连接起来，再逐步延伸连接起全部样点，而形成一条光滑的样条曲线。从力学的角度来看，强迫样条经过某些点，相当于在这些点上对样条施加力的作用，从而使样条发生弯曲，我们截取其中一小段$[x_i, x_{i+1}]$进行分析，并将它置成水平状态。因在该段上有作用力$R_i$，$R_{i+1}$及弯矩$M$，才能使该段样条处于静止的平衡状态，如图 5.9 所示。

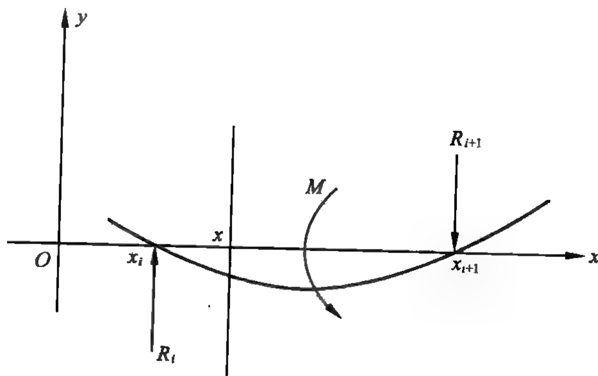



图 5.9

样条在上述力和力矩的作用下发生弯曲变形，其在$x$处的弯曲程度可用曲率半径$R$来衡量。根据材料力学中关于梁的弯曲变形公式

$$\frac{1}{R} = \frac{M(x)}{EJ} \tag{5.179}$$

就可计算出不同$x$处所对应的$y$值(称为挠度)。在公式(5.179)中，$M(x)$是作用于$x$截面上的弯矩，$E$是材料的弹性模数，$J$是$x$截面上材料的质量对截面质心的惯性矩。由于

$$M(x) = R_i(x - x_i) + R_{i+1}(x_{i+1} - x) - M = Ax + B$$

$$\frac{1}{R} = \frac{y''}{[1+(y')^2]^{\frac{3}{2}}} \tag{5.180}$$

当变形小时，$y' \approx 0$，所以$\frac{1}{R} \approx y''$。因此公式(5.179)可近似表为

$$y'' = \frac{M(x)}{EJ} = \frac{Ax+B}{EJ} = Cx + D \tag{5.181}$$

积分上式两次得

$$y = ax^3 + bx^2 + cx + d \tag{5.182}$$

即样条的每一小段曲线可以用一个三次多项式来描述。近年来，样条插值的理论和应用发展很快，它在飞机、船舶、汽车等外形设计的工程问题中，在数值微分、数值积分、微分方程和积分方程的数值解法，以及观测和实验数据处理等方面，样条函数都有其重要的

应用，它已成为现代函数逼近的一个活跃而重要的分支。三次样条是应用最广泛的样条，本节仅介绍三次样条函数，并只用于插值。下面用数学语言来描述，并建立它们的计算公式。

设在区间$[a,b]$上取 $n+1$ 个节点

$$a=x_0<x_1<x_2<\cdots<x_{n-1}<x_n=b$$

给定这些节点上的函数值 $y_i=f(x_i)(i=0,1,2,\cdots,n)$。现要求构造一个分段三次样条函数

$$S(x)=\begin{cases}P_0(x), & [x_0,x_1]\\ P_1(x), & [x_1,x_2]\\ \cdots & \\ P_{n-1}(x), & [x_{n-1},x_n]\end{cases} \tag{5.183}$$

使满足下列条件：

① $S(x_i)=y_i \quad (i=0,1,2,\cdots,n)$；

② 在每个小区间$[x_i,x_{i+1}]$上是一个三次多项式；

③ 在内节点上，$S(x)$具有连续的一阶与二阶导数。

## 8.2 三次样条函数的表达式

我们首先考虑在一个区间$[x_i,x_{i+1}]$上如何决定三次样条函数的问题。因在该区间上满足 $P_i(x_i)=y_i,P_i(x_{i+1})=y_{i+1}$，按牛顿基本差商公式知

$$P_i(x)=P_i(x_i)+(x-x_i)P_i[x_i,x_{i+1}]+(x-x_i)(x-x_{i+1})P_i[x_i,x_{i+1},x] \tag{5.184}$$

其一阶导数及二阶导数为

$$\begin{aligned}P'_i(x)=&P_i[x_i,x_{i+1}]+(2x-x_i-x_{i+1})P_i[x_i,x_{i+1},x]+\\&(x-x_i)(x-x_{i+1})P'_i[x_i,x_{i+1},x]\end{aligned} \tag{5.185}$$

$$\begin{aligned}P''_i(x)=&2P_i[x_i,x_{i+1},x]+2(2x-x_i-x_{i+1})P'_i[x_i,x_{i+1},x]+\\&(x-x_i)(x-x_{i+1})P''_i[x_i,x_{i+1},x]\end{aligned} \tag{5.186}$$

因

$$\begin{cases}P_i(x)——x\text{ 的三次多项式}\\ P'_i(x)——x\text{ 的二次多项式}\\ P''_i(x)——x\text{ 的一次多项式}\end{cases}$$

若已知

$$P''_i(x)=\begin{cases}k_i, & x=x_i\\ k_{i+1}, & x=x_{i+1}\end{cases}$$

则

$$P''_i(x)=k_i+\frac{k_{i+1}-k_i}{x_{i+1}-x_i}(x-x_i) \tag{5.187}$$

利用差商与导数的关系得

$$P'_i[x_i,x_{i+1},x]=P_i[x_i,x_{i+1},x,x]=\frac{P'''_i(\xi)}{3!}=\frac{1}{6}\frac{k_{i+1}-k_i}{x_{i+1}-x_i}$$

$$P''_i[x_i,x_{i+1},x]=0 \tag{5.188}$$

代入式(5.186)后得

$$\begin{aligned}P''_i(x)&=2P_i[x_i,x_{i+1},x]+\frac{2}{6}(2x-x_i-x_{i+1})\frac{k_{i+1}-k_i}{x_{i+1}-x_i}\\&=2P_i[x_i,x_{i+1},x]+\frac{2}{6}[2(x-x_i)+(x_i-x_{i+1})]\frac{k_{i+1}-k_i}{x_{i+1}-x_i}\end{aligned}$$

$$= 2P_i[x_i,x_{i+1},x] + \frac{2}{6}[2\frac{k_{i+1}-k_i}{x_{i+1}-x_i}(x-x_i)+k_i-k_{i+1}]$$

$$= 2P_i[x_i,x_{i+1},x] + \frac{2}{6}[2(P''_i(x)-k_i)+k_i-k_{i+1}]$$

从上式解得

$$P_i[x_i,x_{i+1},x] = \frac{1}{6}[k_i+k_{i+1}+P''_i(x)] \tag{5.189}$$

代入式(5.184)后得区间$[x_i,x_{i+1}]$上的三次样条函数为

$$P_i(x) = y_i+(x-x_i)f[x_i,x_{i+1}]+\frac{1}{6}(x-x_i)(x-x_{i+1})[P''_i(x)+k_i+k_{i+1}] \tag{5.190}$$

而在整个区间$[a,b]$上的三次样条函数则由 $P_0(x),P_1(x),\cdots,P_{n-1}(x)$ 构成。如果已知 $k_0$，$k_1,\cdots,k_n$ 的值，则对于给定的 $x\in[x_i,x_{i+1}]$，便可按公式(5.190)进行插值计算了。因此问题转化为如何确定 $k_0,k_1,\cdots,k_n$ 的问题。

### 8.3　$k_i(i=0,1,2,\cdots,n)$的求解方法

为了求 $k_i(i=0,1,2,\cdots,n)$，首先要推导它们所满足的方程。利用三次样条函数在$[x_0,x_n]$上具有连续的一阶导数条件，要求

$$P'_{i-1}(x_i) = P'_i(x_i) \qquad (i=1,2,\cdots,n-1) \tag{5.191}$$

即对于每个 $x_i(i=1,2,\cdots,n-1)$点处，从它左右两子区间上所决定的三次曲线应该有相同的一阶导数，否则一阶导数在该点就不连续了。由公式(5.185)知

$$\begin{aligned} P'_i(x) &= P_i[x_i,x_{i+1}]+(2x-x_i-x_{i+1})P_i[x_i,x_{i+1},x]+ \\ &\quad (x-x_i)(x-x_{i+1})P'_i[x_i,x_{i+1},x] \\ &= \frac{y_{i+1}-y_i}{x_{i+1}-x_i}+\frac{1}{6}(2x-x_i-x_{i+1})[P''_i(x)+k_i+k_{i+1}]+ \\ &\quad \frac{1}{6}\frac{k_{i+1}-k_i}{x_{i+1}-x_i}(x-x_i)(x-x_{i+1}) \end{aligned} \tag{5.192}$$

$$\begin{aligned} P'_{i-1}(x) &= \frac{y_i-y_{i-1}}{x_i-x_{i-1}}+\frac{1}{6}(2x-x_{i-1}-x_i)[P''_{i-1}(x)+k_{i-1}+k_i]+ \\ &\quad \frac{1}{6}\frac{k_i-k_{i-1}}{x_i-x_{i-1}}(x-x_{i-1})(x-x_i) \end{aligned}$$

代入 $x_i$ 得

$$P'_i(x_i) = \frac{y_{i+1}-y_i}{x_{i+1}-x_i}-\frac{1}{6}(x_{i+1}-x_i)(2k_i+k_{i+1}) \tag{5.193}$$

$$P'_{i-1}(x_i) = \frac{y_i-y_{i-1}}{x_i-x_{i-1}}+\frac{1}{6}(x_i-x_{i-1})(k_{i-1}+2k_i) \tag{5.194}$$

代入式(5.191)便得如下一组方程

$$\frac{y_{i+1}-y_i}{x_{i+1}-x_i}-\frac{1}{6}(x_{i+1}-x_i)(2k_i+k_{i+1}) = \frac{y_i-y_{i-1}}{x_i-x_{i-1}}+\frac{1}{6}(x_i-x_{i-1})(k_{i-1}+2k_i)$$

整理后得

$$(x_i-x_{i-1})k_{i-1}+2(x_{i+1}-x_{i-1})k_i+(x_{i+1}-x_i)k_{i+1}$$

$$= 6\left(\frac{y_{i+1}-y_i}{x_{i+1}-x_i}-\frac{y_i-y_{i-1}}{x_i-x_{i-1}}\right)=6[f[x_i,x_{i+1}]-f[x_{i-1},x_i]]$$

或
$$\frac{x_i-x_{i-1}}{x_{i+1}-x_{i-1}}k_{i-1}+2k_i+\frac{x_{i+1}-x_i}{x_{i+1}-x_{i-1}}k_{i+1}=6\frac{f[x_i,x_{i+1}]-f[x_{i-1},x_i]}{x_{i+1}-x_{i-1}}$$
$$=6f[x_{i-1},x_i,x_{i+1}] \tag{5.195}$$

引入记号

$$\begin{cases} h_i=x_i-x_{i-1},\quad h_{i+1}=x_{i+1}-x_i,\quad h_i+h_{i+1}=x_{i+1}-x_{i-1} \\ \lambda_i=\dfrac{h_{i+1}}{h_i+h_{i+1}}=\dfrac{x_{i+1}-x_i}{x_{i+1}-x_{i-1}},\quad \mu_i=1-\lambda_i=\dfrac{h_i}{h_i+h_{i+1}}=\dfrac{x_i-x_{i-1}}{x_{i+1}-x_{i-1}} \\ d_i=6f[x_{i-1},x_i,x_{i+1}] \end{cases} \tag{5.196}$$

则式(5.195)可化为

$$\mu_i k_{i-1}+2k_i+\lambda_i k_{i+1}=d_i \qquad (i=1,2,\cdots,n-1) \tag{5.197}$$

由于在力学上将 $k_i$ 解释为梁在 $x_i$ 处的弯矩，故式(5.197)称为三弯矩方程组。这个线性方程组有 $n+1$ 个未知量 $k_i(i=0,1,2,\cdots,n)$，所以从这个方程组还求不出解来，因此还需要根据实际情况补充两个边界条件。

① 自然边界条件。这一条件为 $k_0=k_n=0$，即样条在首尾两端自然伸直。这时式(5.197)中的未知量个数与方程数相同，得

$$\begin{cases} 2k_1+\lambda_1k_2=d_1 \\ \mu_ik_{i-1}+2k_i+\lambda_ik_{i+1}=d_i,\quad (i=2,3,\cdots,n-2) \\ \mu_{n-1}k_{n-2}+2k_{n-1}=d_{n-1} \end{cases} \tag{5.198}$$

或写成

$$\begin{bmatrix} 2 & \lambda_1 & & & & \\ \mu_2 & 2 & \lambda_2 & & 0 & \\ & \mu_3 & 2 & \lambda_3 & & \\ & & \ddots & \ddots & \ddots & \\ & 0 & & \mu_{n-2} & 2 & \lambda_{n-2} \\ & & & & \mu_{n-1} & 2 \end{bmatrix}\begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ \cdots \\ k_{n-2} \\ k_{n-1}\end{bmatrix}=\begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ \cdots \\ d_{n-2} \\ d_{n-1}\end{bmatrix} \tag{5.199}$$

② 给定两端斜率 $y'_0,y'_n$。这时在式(5.193)中可令 $i=0$ 得

$$\frac{y_1-y_0}{h_1}-\frac{h_1}{6}(2k_0+k_1)=y'_0$$

解得
$$2k_0+k_1=\frac{6}{h_1}\left(\frac{y_1-y_0}{h_1}-y'_0\right)=d_0 \tag{5.200}$$

同法在式(5.194)中令 $i=n$，解得

$$k_{n-1}+2k_n=\frac{6}{h_n}\left(y'_n-\frac{y_n-y_{n-1}}{h_n}\right)=d_n \tag{5.201}$$

最后把式(5.200)，式(5.197)，式(5.201)联立在一起得 $n+1$ 个方程

$$\mu_ik_{i-1}+2k_i+\lambda_ik_{i+1}=d_i,(i=0,1,2,\cdots,n) \tag{5.202}$$

其中，$\mu_0=0,\lambda_n=0$。亦可用矩阵形式表为

$$\begin{bmatrix} 2 & 1 & & & & \\ \mu_1 & 2 & \lambda_1 & & 0 & \\ & \mu_2 & 2 & \lambda_2 & & \\ & & \ddots & \ddots & \ddots & \\ 0 & & & \mu_{n-1} & 2 & \lambda_{n-1} \\ & & & & 1 & 2 \end{bmatrix} \begin{bmatrix} k_0 \\ k_1 \\ k_2 \\ \cdots \\ k_{n-1} \\ k_n \end{bmatrix} = \begin{bmatrix} d_0 \\ d_1 \\ d_2 \\ \cdots \\ d_{n-1} \\ d_n \end{bmatrix} \tag{5.203}$$

以上三弯矩方程组可用追赶法求解，由于

$$|\lambda_i|+|\mu_i|=\frac{h_{i+1}}{h_i+h_{i+1}}+\frac{h_i}{h_i+h_{i+1}}=1<2$$

即三弯矩方程组的系数矩阵具有强对角线占优特性，因此一定能用消元法求解。

③ 封闭光滑曲线条件 $y_0^{(i)}=y_n^{(i)}(i=0,1,2)$。

根据 $y'_0=y'_n$ 可获得以下方程

$$\frac{y_1-y_0}{h_1}-\frac{h_1}{6}(2k_0+k_1)=\frac{y_n-y_{n-1}}{h_n}+\frac{1}{6}h_n(k_{n-1}+2k_n)$$

再以 $y_n=y_0, y''_n=y''_0$ 代入上式后得

$$\begin{aligned} h_1(2k_0+k_1)+h_n(k_{n-1}+2k_0) &= 6\left[\frac{y_1-y_0}{h_1}-\frac{y_0-y_{n-1}}{h_n}\right] \\ &= 6\{f[x_0,x_1]-f[x_{n-1},x_0]\} \end{aligned}$$

或

$$\mu_n k_{n-1}+2k_0+\lambda_n k_1=d_n \tag{5.204}$$

其中 $\lambda_n=\frac{h_1}{h_n+h_1},\mu_n=\frac{h_n}{h_n+h_1},d_n=6f[x_{n-1},x_0,x_1]$。将式(5.204)与式(5.197)联立便得到以下线性方程组

$$\begin{cases} \mu_i k_{i-1}+2k_i+\lambda_i k_{i+1}=d_i \quad (i=1,2,\cdots,n-2) \\ \mu_{n-1}k_{n-2}+2k_{n-1}+\lambda_{n-1}k_0=d_{n-1} \\ \mu_n k_{n-1}+2k_0+\lambda_n k_1=d_n \end{cases} \tag{5.205}$$

由式(5.205)可解得 $k_0,k_1,\cdots,k_{n-1}$。

三次样条函数 $S(x)$ 逼近 $f(x)$ 是收敛的，并且也是数值稳定的，其误差估计与收敛性证明较复杂，下面只给出结论。

**定理 5.2**　假设函数 $f(x)$ 在 $[a,b]$ 上具有四阶连续导数，$S(x)$ 是 $f(x)$ 满足下述边界条件

①$S'(x_0)=f'_0, S'(x_n)=f'_n$

②$S''(x_0)=f''_0, S''(x_n)=f''_n$

之一的三次样条函数，则有估计式

$$\|f^{(k)}(x)-S^{(k)}(x)\|_\infty \leqslant C_k\|f^{(4)}(x)\|_\infty h^{4-k}, \quad k=0,1,2$$

其中

$$h_i=x_i-x_{i-1}, h=\max_{1\leqslant i\leqslant n} h_i$$

$$C_0=\frac{5}{384}, C_1=\frac{1}{24}, C_2=\frac{1}{8}$$

这个定理告诉我们，当 $h\to 0$（即 $n\to\infty$）时，对于满足第①、②边界条件的三次样条函数 $S(x)$、$S'(x)$ 及 $S''(x)$ 在区间 $[a,b]$ 上分别一致收敛于 $f(x)$、$f'(x)$ 及 $f''(x)$。

**例 5.19** 给出表 5.32。

**表 5.32**

| $x$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $f(x)$ | 0 | 3 | 4 | 6 |
| $f'(x)$ | 1 | | | 0 |

试求 $f(x)$在区间[0,3]上的三次样条函数。

**解** 按表值建立差商表 5.33,由表 5.33 得到 $f[0,1,2]=-1,f[1,2,3]=0.5$。计算 $\lambda_i=\mu_i=\frac{1}{2}(i=1,2)$。

**表 5.33**

| $x$ | $y$ | 一阶差商 | 二阶差商 |
|---|---|---|---|
| 0 | 0 | | |
| | | 3 | |
| 1 | 3 | | −1 |
| | | 1 | |
| 2 | 4 | | 0.5 |
| | | 2 | |
| 3 | 6 | | |

按式(5.202)建立三弯矩方程组

$$\begin{cases} 2k_0+k_1=\dfrac{6}{1-0}(\dfrac{3-0}{1-0}-1)=12 \\ \dfrac{1}{2}k_0+2k_1+\dfrac{1}{2}k_2=6f[0,1,2]=6\times(-1)=-6 \\ \dfrac{1}{2}k_1+2k_2+\dfrac{1}{2}k_3=6f[1,2,3]=6\times0.5=3 \\ k_2+2k_3=\dfrac{6}{3-2}\left(0-\dfrac{6-4}{3-2}\right)=-12 \end{cases}$$

解得 $k_0=2.666\ 67,\quad k_1=6.666\ 67,\quad k_2=5.333\ 33,\quad k_3=-8.666\ 67$

代入公式(5.190)得[0,3]在区间上的三次样条函数 $s(x)$

$$\begin{cases} P_0(x)=3x+\dfrac{1}{6}x(x-1)[P''_0(x)+9.333\ 34] \\ P''_0(x)=2.666\ 67+4x, \qquad 0\leqslant x\leqslant 1 \end{cases}$$

$$\begin{cases} P_1(x)=3+(x-1)+\dfrac{1}{6}(x-1)(x-2)[P''_1(x)+12] \\ P''_1(x)=6.666\ 67-1.333\ 34(x-1), \qquad 1\leqslant x\leqslant 2 \end{cases}$$

$$\begin{cases} P_2(x)=4+2(x-2)+\dfrac{1}{6}(x-2)(x-3)[P''_2(x)-3.333\ 34] \\ P''_2(x)=5.333\ 33-14(x-2), \qquad 2\leqslant x\leqslant 3 \end{cases}$$

## §9 多元函数插值

为叙述简便计,我们只讨论二元函数的插值问题。对于多元函数的插值问题可依照二元

函数的插值问题按同法处理。

## 9.1 直线插值法

设已知双变量函数的数值为

$$z_{ij} = f(x_i, y_j) \qquad (i = 0,1,2,\cdots,n; j = 0,1,2,\cdots,m) \tag{5.206}$$

如表 5.34。当 $x_i<x<x_{i+1}, y_j<y<y_{j+1}$ 时，其对应的 $z$ 值可近似地取为

表 5.34

| $ij$ \ $i$ / $j$ | $x_0$ | $x_1$ | $x_2$ | $\cdots$ | $x_n$ |
|---|---|---|---|---|---|
| $y_0$ | $z_{00}$ | $z_{10}$ | $z_{20}$ | $\cdots$ | $z_{n0}$ |
| $y_1$ | $z_{01}$ | $z_{11}$ | $z_{21}$ | $\cdots$ | $z_{n1}$ |
| $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ |
| $y_m$ | $z_{0m}$ | $z_{1m}$ | $z_{2m}$ | $\cdots$ | $z_{nm}$ |

$$\begin{aligned} z &\approx f(x_i, y_j) + \frac{\partial f(x_i, y_j)}{\partial x}\mathrm{d}x + \frac{\partial f(x_i, y_j)}{\partial y}\mathrm{d}y \\ &\approx f(x_i, y_j) + \frac{f(x_{i+1}, y_j) - f(x_i, y_j)}{x_{i+1} - x_i}(x - x_i) + \\ &\quad \frac{f(x_i, y_{j+1}) - f(x_i, y_j)}{y_{j+1} - y_j}(y - y_j) \\ &= z_{ij} + \frac{z_{(i+1)j} - z_{ij}}{x_{i+1} - x_i}(x - x_i) + \frac{z_{i(j+1)} - z_{ij}}{y_{j+1} - y_j}(y - y_j) \end{aligned} \tag{5.207}$$

## 9.2 曲线插值法

今以实例来说明这种插值方法。

**例 5.20** 已知数值表表 5.35，求 $f(0.5, 0.03)$ 的近似值。

表 5.35

| $z$ \ $x$ / $y$ | 0.4 | 0.7 | 1.0 |
|---|---|---|---|
| 0.00 | 2.500 | 1.429 | 1.000 |
| 0.05 | 2.487 | 1.419 | 0.995 |
| 0.10 | 2.456 | 1.400 | 0.981 |

**解** 按下面的步骤进行插值计算。

① 当 $y_0=0.00$ 时，利用数据表表 5.36 建立差分表表 5.37。

表 5.36

| $x$ | 0.4 | 0.7 | 1.0 |
|---|---|---|---|
| $z$ | 2.500 | 1.429 | 1.00 |

表 5.37

| $x$ | $z$ | $\Delta z$ | $\Delta^2 z$ |
|---|---|---|---|
| 0.4 | 2.500 | | |
| | | −1.071 | |
| 0.7 | 1.429 | | 0.642 |
| | | −0.429 | |
| 1.0 | 1.000 | | |

利用牛顿前插公式计算 $x=0.5$ 时的 $f(0.5,0.00)$的近似值 $z_0$ $(t=(0.5-0.4)/0.3=\frac{1}{3})$

$$z_0=2.500+\frac{1}{3}(-1.071)+\frac{\frac{1}{3}\times\left(-\frac{2}{3}\right)}{2}\times 0.642=2.072$$

② 当 $y_1=0.05$ 时，利用数据表表 5.38 建立差分表表 5.39。

表 5.38

| $x$ | 0.4 | 0.7 | 1.0 |
|---|---|---|---|
| $z$ | 2.487 | 1.419 | 0.995 |

表 5.39

| $x$ | $z$ | $\Delta z$ | $\Delta^2 z$ |
|---|---|---|---|
| 0.4 | 2.487 | | |
| | | −1.068 | |
| 0.7 | 1.419 | | 0.644 |
| | | −0.424 | |
| 1.0 | 0.995 | | |

利用牛顿前插公式计算 $x=0.5$ 时的 $f(0.5,0.05)$的近似值 $z_1$ $(t=\frac{1}{3})$

$$z_1=2.487+\frac{1}{3}\times(-1.068)-\frac{1}{9}\times 0.644=2.069$$

③ 当 $y_2=0.10$ 时，利用数据表表 5.40 建立差分表表 5.41。

表 5.40

| $x$ | 0.4 | 0.7 | 1.0 |
|---|---|---|---|
| $z$ | 2.456 | 1.400 | 0.981 |

表 5.41

| $x$ | $z$ | $\Delta z$ | $\Delta^2 z$ |
|---|---|---|---|
| 0.4 | 2.456 | | |
| | | −1.056 | |
| 0.7 | 1.400 | | 0.637 |
| | | −0.419 | |
| 1.0 | 0.981 | | |

利用牛顿前插公式计算 $x=0.5$ 时的 $f(0.5,0.10)$ 的近似值 $z_2\left(t=\frac{1}{3}\right)$

$$z_2=2.456+\frac{1}{3}\times(-1.056)-\frac{1}{9}\times 0.637=2.033$$

④ 利用以上计算结果得下面数值表表 5.42。

**表 5.42**

| $y$ | 0.00 | 0.05 | 0.10 |
|---|---|---|---|
| $z$ | 2.072 | 2.069 | 2.033 |

按表 5.42 数据建立差分表表 5.43。

**表 5.43**

| $y$ | $z$ | $\Delta z$ | $\Delta^2 z$ |
|---|---|---|---|
| 0.00 | 2.072 | | |
| | | −0.003 | |
| 0.05 | 2.069 | | −0.033 |
| | | −0.036 | |
| 0.10 | 2.033 | | |

取 $y_0=0, t=\frac{0.03-0}{0.05}=\frac{3}{5}$，使用牛顿前插公式计算 $f(0.5,0.03)$ 的近似值

$$f(0.5,0.03)\approx 2.072+\frac{3}{5}\times(-0.003)+\frac{\frac{3}{5}\times\left(-\frac{2}{5}\right)}{2}\times(-0.033)=2.074$$

# 习题五

5.1　已知函数表

| $x$ | 1.45 | 1.36 | 1.14 |
|---|---|---|---|
| $y$ | 3.14 | 4.15 | 5.65 |

用拉格朗日插值公式计算 $x=1.4$ 的函数近似值。

5.2　对于下表

| $x$ | 93.0 | 96.2 | 100.00 | 104.2 | 108.7 |
|---|---|---|---|---|---|
| $y$ | 11.38 | 12.80 | 14.70 | 17.07 | 19.91 |

用牛顿基本差商公式求 $y(102)$ 的近似值。

5.3　对于

$$f(x)=10x^3+x^2,\quad x=0.0,0.1,0.2,0.3,0.4$$

写出差分表。

5.4　给定数表

| $x$ | 75 | 76 | 77 | 78 | 79 | 80 |
|---|---|---|---|---|---|---|
| $y$ | 2.768 | 2.833 | 2.903 | 2.979 | 3.062 | 3.153 |

(1) 作一分段线性插值函数。

(2) 取自然边界条件作三次样条插值多项式。

(3) 用两种插值函数分别计算 $x=75.5$ 和 $x=78.3$ 的函数值。

5.5　已知函数表

| $x$ | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 |
|---|---|---|---|---|---|
| $y$ | 1.000 00 | 1.221 40 | 1.491 82 | 1.822 12 | 2.225 54 |

用牛顿前插公式、斯梯林插值公式、牛顿后插公式分别求 $y(0.05)$、$y(0.42)$、$y(0.75)$ 的近似值。

5.6　按下表构造埃尔米特插值多项式。

(1)

| $x$ | −1 | 0 | 1 |
|---|---|---|---|
| $y$ | −1 | 0 | 1 |
| $y'$ | 0 | 0 | 0 |

(2)

| $x$ | 0 | 1 | 2 |
|---|---|---|---|
| $y$ | 0 | 1 | 1 |
| $y'$ | 0 | 1 | |

5.7　给定数表如下

| $i$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $x_i$ | 0.25 | 0.30 | 0.39 | 0.45 | 0.53 |
| $y_i$ | 0.500 0 | 0.547 7 | 0.624 5 | 0.670 8 | 0.728 0 |

求三次样条函数 $S(x)$，使满足：

(1)端点条件为

$$S'(0.25)=1.0000,\quad S'(0.53)=0.6868$$

(2)端点条件为

$$S''(0.25)=-2,\quad S''(0.53)=0.6479$$

5.8 已知下列数值表

| $x$ | 0.45 | 0.46 | 0.47 | 0.48 | 0.49 | 0.50 |
|---|---|---|---|---|---|---|
| $y$ | 0.475 481 8 | 0.484 655 5 | 0.493 745 2 | 0.502 749 8 | 0.511 668 3 | 0.520 499 9 |

利用反插值的拉格朗日插值公式计算 $y=0.5$ 时的 $x$ 值。

5.9 给定以下数值表

| $x$ | 2.4 | 2.5 | 2.6 |
|---|---|---|---|
| $y$ | 0.002 5 | −0.048 4 | −0.096 8 |

利用插值多项式反插值法求 $y=0$ 对应的 $x$ 值，要求精确到小数后 3 位。

5.10 设拉格朗日插值公式为 $\sum\limits_{i=0}^{n}a_i(x)f(x_i)$。证明：

(1) $\sum\limits_{i=0}^{n}x_i^k a_i(x)\equiv x^k \qquad (k=0,1,2,\cdots,n)$

(2) $\sum\limits_{i=0}^{n}(x_i-x)^k a_i(x)\equiv 0 \qquad (k=0,1,2,\cdots,n)$

5.11 $f(x)=\ln(x)$，插值节点为 $x=0.4,0.5,0.7$ 和 $0.8$，求三次插值多项式 $P_3(x)$，并估计余式。

5.12 对函数 $\sin x,\cos x$ 在 $[0,\frac{\pi}{2}]$ 中构造步长为 $h$ 的等距节点函数表，若要求线性插值的截断误差小于 $0.5\times10^{-8}$，问 $h$ 应取多小及函数表的数值应取至小数后几位较合适？

5.13 给定 $f(x)=e^x$，设 $x=0$ 是四重插值节点，$x=1$ 是单重插值节点，试求相应的埃尔米特插值公式。

5.14 设 $f(x)=x^7+x^4+3x+1$，求下列差商值。

$$f[2^0,2^1,2^2,\cdots,2^7]$$
$$f[2^0,2^1,2^2,\cdots,2^8]$$

5.15 证明 $\sum\limits_{k=0}^{n-1}\Delta^2 y_k=\Delta y_n-\Delta y_0$。

5.16 对于 $f(x)=e^{-x}$ 的下列表值

| $x$ | 0.10 | 0.15 | 0.25 | 0.30 |
|---|---|---|---|---|
| $f(x)$ | $f(0.1)$ | $f(0.15)$ | $f(0.25)$ | $f(0.30)$ |

应用拉格朗日插值公式得

$$L_3(0.20)=-\frac{1}{6}f(0.10)+\frac{2}{3}f(0.15)+\frac{2}{3}f(0.25)-\frac{1}{6}f(0.30)$$

$$R_3(0.20)=0.10\times0.05\times(-0.05)\times(-0.10)\cdot\frac{e^{-\xi}}{4},0.10\leqslant\xi\leqslant0.30$$

问计算中 $f(x)$ 应取几位字长比较合适?

5.17　用线性插值法求 cos50°,表值如下

| $x$ | $\frac{\pi}{4}$ | $\frac{\pi}{3}$ |
|---|---|---|
| $\cos x$ | 0.707 1 | 0.500 0 |

试估计结果的总误差。

5.18　已知某函数有以下表值

| $x$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 | 1.1 | 1.3 |
|---|---|---|---|---|---|---|---|
| $y$ | 0.003 | 0.067 | 0.148 | 0.248 | 0.370 | 0.518 | 0.697 |

如果用代数多项式吻合这些数据,问使用几次多项式为好?

5.19　假设对 $f(x)$ 在步长为 $h$ 的等距节点上造函数表,且 $|f''(x)|\leqslant M$,若取 $f(x)=\sin x$,问 $h$ 应取多大才能保证线性插值的截断误差不大于 $0.5\times10^{-6}$。

# 第六章　数值积分和数值微分

数值积分和数值微分是数值逼近的一个重要内容，也是插值函数的一个直接应用。

## §1　数值积分

在工程和科学实验中，经常要计算定积分

$$\int_a^b f(x)\mathrm{d}x \tag{6.1}$$

在实际应用中常会碰到如下情况。

① 被积函数 $f(x)$ 的原函数不能用初等函数表示，如$\int_0^1 \frac{\sin x}{x}\mathrm{d}x$，$\int_0^1 \mathrm{e}^{-x^2}\mathrm{d}x$ 等。

② $f(x)$的原函数的表达式过于复杂。

③ $f(x)$用表格形式给出。

对于上述情况，都要求建立定积分的近似计算方法。数值积分的基本思想是用简单函数 $P(x)$近似代替被积函数，然后建立如下形式的求积公式

$$\int_a^b f(x)\mathrm{d}x \approx c_0 f(x_0) + c_1 f(x_1) + \cdots + c_n f(x_n) \tag{6.2}$$

公式中的系数值 $c_i$，$x_i(i=0,1,2,\cdots,n)$可以有多种不同的确定方法，相应地就产生了不同的求积公式。例如，可以从以下不同的角度出发来建立求积公式。

① 要求求积公式具有最高的代数精确度。

② 求积公式的余式具有最小的绝对值。

③ 求积公式中的系数绝对值之和为最小以达到抑制误差的作用。

④ 系数相等以便于计算。

在实际问题中，应该根据对积分的具体要求建立或选用所需要的求积公式。

### 1.1　对称的求积公式

若用 $n+1$ 个等距节点下的拉格朗日插值公式表达 $f(x)$时，即

$$f(x) = f(x_0 + ht) = \sum_{i=0}^{n} \frac{(-1)^{n-i} t^{[n+1]}}{i!(n-i)!(t-i)} f(x_i) + R_n(t) \tag{6.3}$$

其中

$$R_n(t) = h^{n+1} \frac{t^{[n+1]}}{(n+1)!} f^{(n+1)}(\xi) \tag{6.4}$$

则

$$I = \int_{x'}^{x''} f(x)\mathrm{d}x = h\int_{t'}^{t''} f(x_0 + ht)\mathrm{d}t = Q + R \tag{6.5}$$

其中

$$Q = h\sum_{i=0}^{n}\left[\int_{t'}^{t''} \frac{(-1)^{n-i} t^{[n+1]}}{i!(n-i)!(t-i)}\mathrm{d}t\right] f(x_i) = h\sum_{i=0}^{n} c_i f(x_i) \tag{6.6}$$

$$R = h\int_{t'}^{t''} R_n(t)\mathrm{d}t \tag{6.7}$$

由于积分值 $Q$ 不仅取决于插值公式的次数 $n$，而且取决于积分区间的上下限 $x'$ 与 $x''$，所以引入下面的记号

$$Q_{np}(q) = h\sum_{i=0}^{n} c_i f(x_i) \tag{6.8}$$

式中，$n$ 为插值多项式的次数；积分区间的长度为 $p$ 个 $h$；$q$ 为积分区间左端点的下标，即 $x'=x_q$。这时积分区间右端点 $x''$ 为

$$x''=x_q+ph=(x_0+qh)+ph=x_0+(p+q)h=x_{p+q}$$

例如 $$I = Q_{31}(1)+\frac{11}{720}h^5 f^{(4)}(\xi) = \frac{h}{24}(-f_0+13f_1+13f_2-f_3)+\frac{11}{720}h^5 f^{(4)}(\xi)$$

式中，$Q_{31}(1)$ 表示从 $x'=x_1$ 到 $x''=x_2$ 对三次插值多项式的求积公式。

### 1.1.1 对称的求积公式

设插值区间为$[x_0, x_n]$，积分区间长度为 $ph$ 且积分区间相对于插值区间的中点 $x_{\frac{n}{2}}$ 对称分布，则积分区间的左端点和右端点的下标分别为

$$t' = \frac{n}{2}-\frac{p}{2} = \frac{n-p}{2}, \quad t'' = \frac{n}{2}+\frac{p}{2} = \frac{n+p}{2} \tag{6.9}$$

其求积公式为 $Q_{np}(\frac{n-p}{2})$，显然它完全由 $n$ 和 $p$ 唯一地确定，所以亦可简记为 $Q_{np}$。它的系数计算公式为

$$c_i = \int_{\frac{n-p}{2}}^{\frac{n+p}{2}} \frac{(-1)^{n-i}t^{[n+1]}}{i!(n-i)!(t-i)}\mathrm{d}t \tag{6.10}$$

系数 $c_i$ 的值对于插值区间中心是对称分布的(参看本节 1.1.2)，即有

$$c_i = c_{n-i} \tag{6.11}$$

由于这种对称性，称这种类型的求积公式为对称的求积公式。它的表达式 $Q_{np}$ 亦可改写成对称的形式(这时把区间的中点取为 $O$ 点)

$$Q_{np} = \frac{N}{D}h\sum_{i=-\frac{n}{2}}^{\frac{n}{2}} w_i f(x_i) \tag{6.12}$$

式中，$N$、$D$ 为常数，$w_i=w_{-i}$。为了使用方便，对 $n$=偶数时的对称求积公式

$$Q_{np} = \frac{N}{D}h\sum_{i=-\frac{n}{2}}^{\frac{n}{2}} w_i f(x_i)$$

$$R = \frac{N'}{D'}h^{n+3} f^{(n+2)}(\xi) \tag{6.13}$$

及 $n$=奇数时的对称求积公式

$$Q_{np} = \frac{N}{D}h\sum_{i=-\frac{n}{2}}^{\frac{n}{2}} w_i f(x_i)$$

$$R = \frac{N'}{D'}h^{n+2} f^{(n+1)}(\xi) \tag{6.14}$$

分别列出表 6.1 和表 6.2，以备查用。

表 6.1 （$n$ 为偶数）

| 公式 | $N$ | $D$ | $w_0$ | $w_{\pm1}$ | $w_{\pm2}$ | $w_{\pm3}$ | $N'$ | $D'$ |
|---|---|---|---|---|---|---|---|---|
| $Q_{01}$ | 1 | 1 | 1 | | | | 1 | 24 |
| $Q_{02}$ | 1 | 1 | 2 | | | | 1 | 3 |
| $Q_{21}$ | 1 | 24 | 22 | 1 | | | −17 | 567 |
| $Q_{22}$ | 1 | 3 | 4 | 1 | | | −1 | 90 |
| $Q_{24}$ | 4 | 3 | −1 | 2 | | | 14 | 45 |
| $Q_{41}$ | 1 | 5 760 | 5 178 | 308 | 17 | | 367 | 967 680 |
| $Q_{42}$ | 1 | 90 | 114 | 34 | −1 | | 1 | 756 |
| $Q_{44}$ | 2 | 45 | 12 | 32 | 7 | | −8 | 945 |
| $Q_{46}$ | 3 | 10 | 26 | −14 | 11 | | 41 | 140 |
| $Q_{61}$ | 1 | 967 680 | 862 564 | 57 249 | −5 058 | 367 | 27 859 | 464 486 400 |
| $Q_{62}$ | 1 | 3 780 | 4 688 | 150 3 | −72 | 5 | −23 | 113 400 |
| $Q_{64}$ | 2 | 945 | 332 | 612 | 171 | −4 | 13 | 14 175 |
| $Q_{66}$ | 1 | 140 | 272 | 27 | 216 | 41 | −9 | 1 400 |
| $Q_{68}$ | 8 | 945 | −2 459 | 2 196 | −954 | 460 | 3 956 | 14 175 |

表 6.2 （$n$ 为奇数）

| 公式 | $N$ | $D$ | $w_{\pm\frac{1}{2}}$ | $w_{\pm\frac{3}{2}}$ | $w_{\pm\frac{5}{2}}$ | $w_{\pm\frac{7}{2}}$ | $N'$ | $D'$ |
|---|---|---|---|---|---|---|---|---|
| $Q_{11}$ | 1 | 2 | 1 | | | | −1 | 12 |
| $Q_{13}$ | 3 | 2 | 1 | | | | 3 | 4 |
| $Q_{31}$ | 1 | 24 | 13 | −1 | | | 11 | 720 |
| $Q_{33}$ | 3 | 8 | 3 | 1 | | | −3 | 80 |
| $Q_{35}$ | 5 | 24 | 1 | 11 | | | 95 | 144 |
| $Q_{51}$ | 1 | 1 440 | 802 | −93 | 11 | | −191 | 60 480 |
| $Q_{53}$ | 3 | 160 | 58 | 23 | −1 | | 13 | 2 240 |
| $Q_{55}$ | 5 | 288 | 50 | 75 | 19 | | −275 | 12 096 |
| $Q_{57}$ | 7 | 144 0 | 562 | −453 | 611 | | 525 7 | 864 0 |
| $Q_{71}$ | 1 | 120 960 | 683 23 | −953 1 | 187 9 | −191 | 249 7 | 362 880 0 |
| $Q_{73}$ | 1 | 448 0 | 480 7 | 204 9 | −149 | 13 | −39 | 448 00 |
| $Q_{75}$ | 5 | 241 92 | 447 5 | 580 5 | 187 1 | −55 | 425 | 145 152 |
| $Q_{77}$ | 7 | 172 80 | 298 9 | 132 3 | 357 7 | 751 | −818 3 | 518 400 |
| $Q_{79}$ | 9 | 448 0 | −171 1 | 496 7 | −280 3 | 178 7 | 254 13 | 448 00 |

### 1.1.2 关于 $c_i=c_{n-i}$ 的证明

分四种情况讨论之。

(1) $n=2r$（偶数），$i=2m$（偶数）

由式(6.10)可得

$$c_i=\int_{\frac{n-p}{2}}^{\frac{n+p}{2}}\frac{(-1)^{2(r-m)}t^{[2r+1]}}{i!(n-i)!(t-i)}\mathrm{d}t=\frac{1}{i!(n-i)!}\int_{\frac{n-p}{2}}^{\frac{n+p}{2}}\frac{t^{[2r+1]}}{t-i}\mathrm{d}t$$

令 $\xi=t-\frac{n}{2}$，则 $t=\xi+\frac{n}{2}$，$\xi'=-\frac{p}{2}$，$\xi''=\frac{p}{2}$，代入上式得

$$c_i=\frac{1}{i!(n-i)!}\int_{-\frac{p}{2}}^{\frac{p}{2}}\frac{(\xi+\frac{n}{2})^{[2r+1]}}{\xi+(\frac{n}{2}-i)}\mathrm{d}\xi$$

$$=\frac{1}{i!(n-i)!}\int_{-\frac{p}{2}}^{\frac{p}{2}}\frac{\overbrace{\left[\xi+\frac{n}{2}\right]\cdots\left[\xi+(\frac{n}{2}-i)\right]\cdots}^{r\text{个因子}}[\xi-0]\overbrace{\cdots\left[\xi-(\frac{n}{2}-i)\right]\cdots\left[\xi-\frac{n}{2}\right]}^{r\text{个因子}}}{\xi+(\frac{n}{2}-i)}\mathrm{d}\xi$$

$$=\frac{1}{i!(n-i)!}\int_{-\frac{p}{2}}^{\frac{p}{2}}\varphi(\xi)\left[\xi-(\frac{n}{2}-i)\right]\mathrm{d}\xi \tag{6.15}$$

其中

$$\varphi(\xi)=\left[\xi^2-(\frac{n}{2})^2\right]\cdots\left[\xi^2-\left(\frac{n}{2}-\overline{i-1}\right)^2\right]\left[\xi^2-\left(\frac{n}{2}-\overline{i+1}\right)^2\right]\cdots\left[\xi^2-1^2\right]\cdot\xi \tag{6.16}$$

显然 $\varphi(\xi)$ 是奇函数，所以有

$$\int_{-\frac{p}{2}}^{\frac{p}{2}}\varphi(\xi)\mathrm{d}\xi=0$$

于是式(6.15)可化为

$$c_i=\frac{1}{i!(n-i)!}\left[\int_{-\frac{p}{2}}^{\frac{p}{2}}\varphi(\xi)\xi\mathrm{d}\xi-(\frac{n}{2}-i)\int_{-\frac{p}{2}}^{\frac{p}{2}}\varphi(\xi)\mathrm{d}\xi\right]$$

$$=\frac{1}{i!(n-i)!}\int_{-\frac{p}{2}}^{\frac{p}{2}}\varphi(\xi)\xi\mathrm{d}\xi \tag{6.17}$$

再以 $n-i$ 代替式(6.10)中的 $i$ 值，得

$$c_{n-i}=\int_{\frac{n-p}{2}}^{\frac{n+p}{2}}\frac{(-1)^{n-(n-i)}t^{[n+1]}}{(n-i)![n-(n-i)]![t-(n-i)]}\mathrm{d}t$$

$$=\frac{(-1)^{2m}}{i!(n-i)!}\int_{\frac{n-p}{2}}^{\frac{n+p}{2}}\frac{t^{[2r+1]}}{t-n+i}\mathrm{d}t$$

$$=\frac{1}{i!(n-i)!}\int_{-\frac{p}{2}}^{\frac{p}{2}}\frac{(\xi+\frac{n}{2})^{[2r+1]}}{\xi-(\frac{n}{2}-i)}\mathrm{d}\xi$$

$$=\frac{1}{i!(n-i)!}\int_{-\frac{p}{2}}^{\frac{p}{2}}\varphi(\xi)\left[\xi+(\frac{n}{2}-i)\right]\mathrm{d}\xi$$

$$=\frac{1}{i!(n-i)!}\left[\int_{-\frac{p}{2}}^{\frac{p}{2}}\varphi(\xi)\xi\mathrm{d}\xi+(\frac{n}{2}-i)\int_{-\frac{p}{2}}^{\frac{p}{2}}\varphi(\xi)\mathrm{d}\xi\right]$$

$$=\frac{1}{i!(n-i)!}\int_{-\frac{p}{2}}^{\frac{p}{2}}\varphi(\xi)\xi\mathrm{d}\xi \tag{6.18}$$

对照式(6.17)与式(6.18)知， $c_i=c_{n-i}$成立。

(2) $n=2r$(偶数)， $i=2m+1$(奇数)

可仿(1)证得 $c_i=c_{n-i}$成立。

(3) $n=2r+1$(奇数)， $i=2m$(偶数)

按式(6.10)有

$$c_i=\int_{\frac{n-p}{2}}^{\frac{n+p}{2}}\frac{(-1)^{(2r+1)-2m}t^{[2r+2]}}{i!(n-i)!(t-i)}\mathrm{d}t$$

$$=\frac{(-1)^{2(r-m)+1}}{i!(n-i)!}\int_{\frac{n-p}{2}}^{\frac{n+p}{2}}\frac{t^{[2r+2]}}{t-i}\mathrm{d}t$$

$$=\frac{-1}{i!(n-i)!}\int_{-\frac{p}{2}}^{\frac{p}{2}}\frac{(\xi+\frac{n}{2})^{[2r+2]}}{\xi+(\frac{n}{2}-i)}\mathrm{d}\xi \quad (\text{作变换 }\xi=t-\frac{n}{2})$$

$$=\frac{-1}{i!(n-i)!}\int_{-\frac{p}{2}}^{\frac{p}{2}}\frac{\overbrace{\left[\xi+\frac{n}{2}\right]\cdots\left[\xi+(\frac{n}{2}-i)\right]\cdots\left[\xi+\frac{1}{2}\right]}^{r+1\text{个因子}}\overbrace{\left[\xi-\frac{1}{2}\right]\cdots\left[\xi-(\frac{n}{2}-i)\right]\cdots\left[\xi-\frac{n}{2}\right]}^{r+1\text{个因子}}}{\xi+(\frac{n}{2}-i)}\mathrm{d}\xi$$

$$=\frac{-1}{i!(n-i)!}\int_{-\frac{p}{2}}^{\frac{p}{2}}\varphi(\xi)\left[\xi-(\frac{n}{2}-i)\right]\mathrm{d}\xi \tag{6.19}$$

其中

$$\varphi(\xi)=\left[\xi^2-(\frac{n}{2})^2\right]\cdots\left[\xi^2-(\frac{n}{2}-\overline{i-1})^2\right]\left[\xi^2-(\frac{n}{2}-\overline{i+1})^2\right]\cdots\left[\xi^2-(\frac{1}{2})^2\right]$$

显然 $\varphi(\xi)$为偶函数,所以有

$$c_i=\frac{-1}{i!(n-i)!}\left[\int_{-\frac{p}{2}}^{\frac{p}{2}}\varphi(\xi)\xi\mathrm{d}\xi-(\frac{n}{2}-i)_n\int_{-\frac{p}{2}}^{\frac{p}{2}}\varphi(\xi)\mathrm{d}\xi\right]$$

$$=\frac{\frac{n}{2}-i}{i!(n-i)!}\int_{-\frac{p}{2}}^{\frac{p}{2}}\varphi(\xi)\mathrm{d}\xi \tag{6.20}$$

相应地有

$$c_{n-i}=\int_{\frac{n-p}{2}}^{\frac{n+p}{2}}\frac{(-1)^{n-(n-i)}t^{[n+1]}}{(n-i)![n-(n-i)]![t-(n-i)]}\mathrm{d}t$$

$$=\frac{(-1)^i}{i!(n-i)!}\int_{\frac{n-p}{2}}^{\frac{n+p}{2}}\frac{t^{[n+1]}}{t-n+i}\mathrm{d}t$$

$$=\frac{(-1)^{2m}}{i!(n-i)!}\int_{-\frac{p}{2}}^{\frac{p}{2}}\frac{(\xi+\frac{n}{2})^{[2r+2]}}{\xi+\frac{n}{2}-n+i}\mathrm{d}\xi$$

$$=\frac{1}{i!(n-i)!}\int_{-\frac{p}{2}}^{\frac{p}{2}}\frac{(\xi+\frac{n}{2})^{[2r+2]}}{\xi-(\frac{n}{2}-i)}\mathrm{d}\xi$$

$$= \frac{1}{i!(n-i)!}\int_{-\frac{p}{2}}^{\frac{p}{2}} \varphi(\xi)\left[\xi + \left(\frac{n}{2} - i\right)\right] \mathrm{d}\xi$$

$$= \frac{1}{i!(n-i)!}\left[\int_{-\frac{p}{2}}^{\frac{p}{2}} \varphi(\xi)\xi \mathrm{d}\xi + \left(\frac{n}{2} - i\right)\int_{-\frac{p}{2}}^{\frac{p}{2}} \varphi(\xi) \mathrm{d}\xi\right]$$

$$= \frac{\frac{n}{2} - i}{i!(n-i)!}\int_{-\frac{p}{2}}^{\frac{p}{2}} \varphi(\xi) \mathrm{d}\xi \tag{6.21}$$

对照式(6.20)与式(6.21)可知，$c_i = c_{n-i}$成立。

(4) $n = 2r - 1$(奇数)，　$i = 2m + 1$(奇数)

可仿(3)证得 $c_i = c_{n-i}$成立。

### 1.2　牛顿-柯特斯求积公式

当积分区间与插值区间相同时，相应的对称求积公式为 $Q_{11}, Q_{22}, \cdots, Q_{m}, \cdots$，这些公式统称为牛顿-柯特斯求积公式。其中常用的有

$$Q_{11} = \frac{h}{2}(f_{-\frac{1}{2}} + f_{\frac{1}{2}}),\ R = -\frac{1}{12}h^3 f^{(2)}(\xi) \text{—— 梯形公式} \tag{6.22}$$

$$Q_{22} = \frac{h}{3}(f_{-1} + 4f_0 + f_1),\ R = -\frac{1}{90}h^5 f^{(4)}(\xi) \text{—— 辛卜生公式} \tag{6.23}$$

$$Q_{33} = \frac{3}{8}h(f_{-\frac{3}{2}} + 3f_{-\frac{1}{2}} + 3f_{\frac{1}{2}} + f_{\frac{3}{2}}),\ R = \frac{-3}{80}h^5 f^{(4)}(\xi) \text{——3/8 公式} \tag{6.24}$$

$$Q_{44} = \frac{2}{45}h(7f_{-2} + 32f_{-1} + 12f_0 + 32f_1 + 7f_2), R = \frac{-8}{945}h^7 f^{(6)}(\xi) \text{—— 柯特斯公式} \tag{6.25}$$

为了讨论舍入误差的影响，可将牛顿-柯特斯求积公式改写为如下形式

$$Q_{nn} = \frac{N}{D}h\sum_{i=-\frac{n}{2}}^{\frac{n}{2}} w_i f_i = nh\sum_{i=-\frac{n}{2}}^{\frac{n}{2}}\left(\frac{Nw_i}{nD}\right)f_i = (b-a)\sum_{i=-\frac{n}{2}}^{\frac{n}{2}} c_i^{(n)} f_i \tag{6.26}$$

式中，$c_i^{(n)}$ 叫做牛顿-柯特斯系数，它与积分区间无关，其部分系数数值列于表 6.3 中。

当 $f(x) = c$(常数)时有

$$\int_a^b c\,\mathrm{d}x = (b-a)\sum_{i=-\frac{n}{2}}^{\frac{n}{2}} c_i^{(n)} f_i + R$$

$$c(b-a) = (b-a)c\sum_{i=-\frac{n}{2}}^{\frac{n}{2}} c_i^{(n)} + 0$$

即

$$\sum_{i=-\frac{n}{2}}^{\frac{n}{2}} c_i^{(n)} \equiv 1 \tag{6.27}$$

成立。式(6.27)不仅可以用做检验牛顿-柯特斯求积公式中系数 $c_i^{(n)}$ 的计算正确性；还可以用来估计计算结果 $Q_{nn}$的舍入误差大小。

表 6.3

| $n$ | $c_0^{(n)}$ | $c_1^{(n)}$ | $c_2^{(n)}$ | $c_3^{(n)}$ | $c_4^{(n)}$ | $c_5^{(n)}$ | $c_6^{(n)}$ | $c_7^{(n)}$ | $c_8^{(n)}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | $\frac{1}{2}$ | $\frac{1}{2}$ | | | | | | | |
| 2 | $\frac{1}{6}$ | $\frac{4}{6}$ | $\frac{1}{6}$ | | | | | | |
| 3 | $\frac{1}{8}$ | $\frac{3}{8}$ | $\frac{3}{8}$ | $\frac{1}{8}$ | | | | | |
| 4 | $\frac{7}{90}$ | $\frac{16}{45}$ | $\frac{2}{15}$ | $\frac{16}{45}$ | $\frac{7}{90}$ | | | | |
| 5 | $\frac{19}{288}$ | $\frac{25}{96}$ | $\frac{25}{144}$ | $\frac{25}{144}$ | $\frac{25}{96}$ | $\frac{19}{288}$ | | | |
| 6 | $\frac{41}{840}$ | $\frac{9}{35}$ | $\frac{9}{280}$ | $\frac{34}{105}$ | $\frac{9}{280}$ | $\frac{9}{35}$ | $\frac{41}{840}$ | | |
| 7 | $\frac{751}{172\,80}$ | $\frac{357\,7}{172\,80}$ | $\frac{132\,3}{172\,80}$ | $\frac{298\,9}{172\,80}$ | $\frac{298\,9}{172\,80}$ | $\frac{132\,3}{172\,80}$ | $\frac{357\,7}{172\,80}$ | $\frac{751}{172\,80}$ | |
| 8 | $\frac{989}{283\,50}$ | $\frac{588\,8}{283\,50}$ | $\frac{-928}{283\,50}$ | $\frac{104\,96}{283\,50}$ | $\frac{-454\,0}{283\,50}$ | $\frac{104\,96}{283\,50}$ | $\frac{-928}{283\,50}$ | $\frac{588\,8}{283\,50}$ | $\frac{989}{283\,50}$ |

假如 $f(x_i)$和 $c_i^{(n)}$ 的舍入误差$\leqslant$e,那么使用牛顿-柯特斯公式计算积分近似值时引入的舍入误差小于或等于

$$(b-a)\mathrm{e}\sum_{i=0}^{n}\left(\left|c_i^{(n)}\right|+\left|f(x_i)\right|\right) \tag{6.28}$$

由于

$$\sum_{i=0}^{n}\left|c_i^{(n)}\right|\begin{cases}=1, & c_i^{(n)}>0\\ >1, & c_i^{(n)}\gtrless 0\end{cases} \tag{6.29}$$

因此,当牛顿-柯特斯公式中的系数有正有负时,其结果的舍入误差限较系数恒为正时要大。由表 6.3 见,当 $n\geqslant 8$ 时,有的公式中有负系数出现,所以采用牛顿-柯特斯公式时,一般只用 $n\leqslant 7$ 的公式。

**例 6.1**　用 $n=6$ 牛顿-柯特斯公式计算下列定积分值

$$\int_0^1 \frac{\mathrm{d}x}{1+x}$$

**解**　计算出 $h=\frac{1-0}{6}=\frac{1}{6}$时在 $x_i=0+\frac{1}{6}i=\frac{1}{6}i(i=0,1,2,\cdots,6)$上的被积函数值 $f(x_i)=\frac{6}{6+i}$,按 $n=6$ 的牛顿-柯特斯公式得

$$\begin{aligned} I&=\frac{1-0}{840}\left[41\times\frac{6}{6}+216\times\frac{6}{7}+27\times\frac{6}{8}+272\times\frac{6}{9}+27\times\frac{6}{10}+216\times\frac{6}{11}+41\times\frac{6}{12}\right]\\ &=\frac{1}{840}\times 582.244\,372=0.693\,1\end{aligned}$$

## 1.3　复化求积公式

由于上面所讲的原因,高阶的牛顿-柯特斯公式不宜使用。为了提高计算积分的精度,可

以把积分区间$[a, b]$等分为$M$个小段，然后在每$n$个小段上应用$Q_{mn}$公式求出积分值，并累加得$[a, b]$区间上的积分总值。按这种方法建立起来的求积公式称为复化求积公式。

设每$n$个小段为一个大段，记$m$为大段总数，则有如下关系式

$$m = \frac{M}{n},\ h = \frac{b-a}{M} = \frac{b-a}{mn} \tag{6.30}$$

下面介绍几个常用的复化求积公式。

(1) 复化梯形公式

将$[a, b]M$等分得$h=\frac{b-a}{M}$，对$[x_0, x_1]$, $[x_1, x_2]$, …, $[x_{M-1}, x_M]$(其中$x_0=a$, $x_M=b$)的每个区间，应用梯形公式并累加得

$$\int_b^a f(x)\mathrm{d}x = \frac{h}{2}(f_0+f_1)+\frac{h}{2}(f_1+f_2)+\cdots+\frac{h}{2}(f_{M-1}+f_M)+R$$

$$= h(\frac{1}{2}f_0+f_1+\cdots+f_{M-1}+\frac{1}{2}f_M)+R \tag{6.31}$$

其中 $$R=-\frac{Mh^3}{12}f''(\xi)=-\frac{(b-a)h^2}{12}f''(\xi)=-\frac{(b-a)^3}{12M^2}f''(\xi),\ \xi\in(a,b) \tag{6.32}$$

当要求截断误差为$\varepsilon$时，$M$只需满足

$$\frac{(b-a)^3}{12M^2}m_2<\varepsilon \qquad (m_2=\max_{x\in[a,b]}|f''(x)|)$$

或 $$M\geqslant\left[\sqrt{\frac{(b-a)^3m_2}{12\varepsilon}}\right]+1,\text{([ ]为取整值运算符)} \tag{6.33}$$

(2) 复化辛卜生公式

设$M=2m$，则$h=\frac{b-a}{2m}=\frac{b-a}{M}$，对$[x_0, x_2]$, $[x_2, x_4]$, …, $[x_{M-2}, x_M]$中的每个区间上应用辛卜生公式并累加得

$$\int_a^b f(x)\mathrm{d}x = \frac{h}{3}(f_0+4f_1+f_2)+\frac{h}{3}(f_2+4f_3+f_4)+\cdots+\frac{h}{3}(f_{M-2}+4f_{M-1}+f_M)+R$$

$$= \frac{h}{3}[(f_0+f_M)+4(f_1+f_3+\cdots+f_{M-1})+2(f_2+f_4+\cdots+f_{M-2})]+R \tag{6.34}$$

其中 $$R=-\frac{(b-a)^5}{2\,880m^4}f^{(4)}(\xi)=-\frac{b-a}{180}h^4f^{(4)}(\xi),\ \xi\in(a,b) \tag{6.35}$$

当要求截断误差为$\varepsilon$时，只需

$$\frac{(b-a)^5}{2\,880m^4}m_4<\varepsilon,\quad (m_4=\max_{x\in[a,b]}|f^{(4)}(x)|)$$

或 $$m\geqslant\left[\sqrt[4]{\frac{(b-a)^5m_4}{2\,880\varepsilon}}\right]+1 \tag{6.36}$$

**例 6.2** 对定积分$\int_0^1\frac{\sin x}{x}\mathrm{d}x$, $\varepsilon=10^{-6}$，分别应用复化梯形公式和复化辛卜生公式计算时，需$M$取多少合适？

**解** 为了确定$M$，先估计$m_2$, $m_4$，由

$$f(x)=\frac{\sin x}{x}=\int_0^1\frac{\cos tx\,\mathrm{d}tx}{x}=\frac{\sin tx}{x}\bigg|_0^1=\frac{\sin x}{x}$$

从而 $$f'(x)=-\int_0^1 t\sin tx\,dt,\ f''(x)=-\int_0^1 t^2\cos tx\,dt$$

一般 $$f^{(k)}(x)=\int_0^1 t^k\cos\left(tx+\frac{k\pi}{2}\right)dt$$

于是有 $$\left|f^{(k)}(x)\right|\leqslant\int_0^1 t^k\cos\left(tx+\frac{k\pi}{2}\right)dt<\frac{1}{k+1}$$

所以 $m_2<\frac{1}{3}$， $m_4<\frac{1}{5}$。

对复化梯形公式由式(6.33)得

$$M\geqslant\left[\sqrt{\frac{10^6}{36}}\right]+1=167$$

而对复化辛卜生公式，由式(6.36)得

$$m\geqslant\left[\sqrt[4]{\frac{10^6}{2\,880\times5}}\right]+1=3,\quad M=2m=6$$

从本例可见，为达到相同的精度水平，使用复化梯形公式所需的计算量比使用复化辛卜生公式的计算量大。

(3) 复化 3/8 公式

设 $M=3m$，则 $h=\frac{b-a}{3m}=\frac{b-a}{M}$，对$[x_0,x_3]$，$[x_3,x_6]$，…，$[x_{M-3},x_M]$中的每个区间上应用3/8 公式并累加得

$$\int_a^b f(x)dx=\frac{3h}{8}[(f_0+f_M)+2(f_3+f_6+\cdots+f_{M-3})+3(f_1+f_2+f_4+f_5+\cdots+f_{M-2}+f_{M-1})]+R \tag{6.37}$$

其中 $$R=-\frac{3h^5}{80}mf^{(4)}(\xi)=-\frac{3h^5}{80}\frac{M}{3}f^{(4)}(\xi)=-\frac{(b-a)h^4}{80}f^{(4)}(\xi)$$
$$=-\frac{(b-a)^5}{80M^4}f^{(4)}(\xi),\ \xi\in(a,b) \tag{6.38}$$

(4) 复化柯特斯公式

设 $M=4m$，则 $h=\frac{b-a}{4m}=\frac{b-a}{M}$，对$[x_0,x_4]$，$[x_4,x_8]$，…，$[x_{M-4},x_M]$中的每个区间上应用柯特斯公式并累加得

$$\int_a^b f(x)dx=4h\left[\frac{7}{90}f_0+\frac{32}{90}\sum_{i=1}^m f_{4i-3}+\frac{12}{90}\sum_{i=1}^m f_{4i-2}+\frac{32}{90}\sum_{i=1}^m f_{4i-1}+\frac{14}{90}\sum_{i=1}^{m-1} f_{4i}+\frac{7}{90}f_M\right]+R \tag{6.39}$$

其中 $$R=-\frac{2(b-a)h^6}{945}f^{(6)}(\xi),\qquad \xi\in(a,b) \tag{6.40}$$

**例 6.3** 利用复化辛卜生公式计算积分

$$I=\int_0^1\frac{dx}{1+x}$$

**解** 如取 $2m=10$，则 $h=\frac{1-0}{10}=0.1$，按式(6.34)计算

$$I \approx \frac{0.1}{3}\left[\left(\frac{1}{1+0}+\frac{1}{1+1}\right)+4\left(\frac{1}{1+0.1}+\frac{1}{1+0.3}+\frac{1}{1+0.5}+\frac{1}{1+0.7}+\frac{1}{1+0.9}\right)+\right.$$
$$\left.2\left(\frac{1}{1+0.2}+\frac{1}{1+0.4}+\frac{1}{1+0.6}+\frac{1}{1+0.8}\right)\right]$$
$$=0.033\,33[(1+0.5)+4(0.909\,09+0.769\,23+0.666\,67+0.588\,24+0.526\,32)+$$
$$2(0.833\,33+0.714\,29+0.625\,00+0.555\,56)]$$
$$=0.033\,33\times 20.794\,5=0.693\,15 \tag{6.41}$$

其截断误差可估计如下

$$f^{(4)}(x)=\frac{24}{(1+x)^5},\qquad \max_{0\leqslant x\leqslant 1}\left|f^{(4)}(x)\right|=24$$
$$|R|\leqslant\frac{(1-0)\times(0.1)^4}{180}\times 24=1.3\times 10^{-5}$$

为估算舍入误差，设 $f_i$ 的舍入误差为 $\varepsilon_i$，则 $|\varepsilon_i|\leqslant 0.5\times 10^{-5}$。在式(6.41)中数 20.794 5 的舍入误差为

$$|\varepsilon_0+\varepsilon_{10}+4(\varepsilon_1+\varepsilon_3+\varepsilon_5+\varepsilon_7+\varepsilon_9)+2(\varepsilon_2+\varepsilon_4+\varepsilon_6+\varepsilon_8)|\leqslant 30\times 0.5\times 10^{-5}$$

由于 0.033 33 的舍入误差小于 $0.5\times 10^{-5}$，因此结果的舍入误差

$$|\in|\leqslant 0.033\,33\times(30\times 0.5\times 10^{-5})+20.794\,5\times(0.5\times 10^{-5})=0.1\times 10^{-3}$$

结果的总误差为

$$|\varepsilon|\leqslant 1.3\times 10^{-5}+0.1\times 10^{-3}\approx 0.1\times 10^{-3}<0.5\times 10^{-3}$$

取 $I\approx 0.693$。

计算实践表明，利用复化求积公式进行计算，易于编制程序。当需要加密分点以提高精度时，已算出的函数值及积分值仍旧是有用的。以复化梯形公式为例，用 $T_1$ 表示在区间 $[a, b]$ 上使用梯形公式的结果，即

$$T_1=\frac{b-a}{2}[f(a)+f(b)] \tag{6.42}$$

如把区间 $[a, b]$ 分成两个相等的子区间 $[a, a+\frac{b-a}{2}]$ 和 $[a+\frac{b-a}{2}, b]$，这时步长 $h=\frac{b-a}{2}$，使用 $n=2$ 时的复化梯形公式得

$$T_2=\frac{b-a}{2}\left\{\frac{1}{2}[f(a)+f(b)]+f\left(a+\frac{b-a}{2}\right)\right\}$$
$$=\frac{T_1}{2}+\frac{b-a}{2}f\left(a+\frac{b-a}{2}\right) \tag{6.43}$$

继续将 $[a, b]$ 四等分为四个子区间，步长 $h=\frac{b-a}{4}$，分点为

$$x_i=a+\frac{b-a}{4}i\quad(i=0, 1, 2, 3, 4)$$

应用 $n=2^2$ 时的复化梯形公式得

$$T_{2^2}=\frac{T_2}{2}+\frac{b-a}{4}\left[f\left(a+\frac{b-a}{4}\right)+f\left(a+\frac{b-a}{4}\times 3\right)\right] \tag{6.44}$$

类似地可以算出 $T_{2^3}, T_{2^4}, \cdots$。

一般情况，当区间 $[a, b]$ 分为 $2^k$ 等分，步长 $h=\frac{b-a}{2^k}$，则有

$$T_{2^k}=\frac{1}{2}T_{2^{k-1}}+h\sum_{i=1}^{2^{k-1}}f(a+(2i-1)h) \tag{6.45}$$

此即逐次加密分点时的复化梯形递推公式。

为了确定区间$[a, b]$的分段数 $m$，就需要根据余式作估算，但要对余式作精确的估计一般是不容易的。实际计算时可采用变步长的求积方法，通常将步长逐次折半或加倍，反复利用复化求积公式进行计算，直到相邻两次积分近似值相当接近时为止。以复化梯形公式为例，可先算出 $M$ 等分时的积分值 $T_M$，然后计算 $2M$ 等分时的积分值 $T_{2M}$。记

$$I=\int_a^b f(x)\mathrm{d}x$$

则

$$I-T_M=-\frac{b-a}{12}h^2f^{(2)}(\xi_1)$$

$$I-T_{2M}=-\frac{b-a}{12}\left(\frac{h}{2}\right)^2f^{(2)}(\xi_2)$$

假定 $f^{(2)}(x)$在$[a, b]$上变化不大，即 $f^{(2)}(\xi_1)\approx f^{(2)}(\xi_2)$，于是有

$$\frac{I-T_M}{I-T_{2M}}\approx 4$$

或

$$I\approx T_{2M}+\frac{1}{3}(T_{2M}-T_M) \tag{6.46}$$

记

$$\Delta=\frac{1}{3}(T_{2M}-T_M) \tag{6.47}$$

式(6.46)说明用 $T_{2M}$作为 $I$ 的近似值时，它的误差近似于 $\Delta$ 值。故当允许误差为 $\varepsilon$ 时，可建立如下的变步长积分方法：当$|\Delta|>\varepsilon$ 时，反复将 $h$ 折半进行计算直到$|\Delta|<\varepsilon$ 即可，取最后的 $T_{2M}$为结果；当$|\Delta|<\varepsilon$ 时，反复将步长 $h$ 加倍进行计算直到$|\Delta|>\varepsilon$ 即可，这时再将步长折半一次进行计算，就得到所要的结果。

对复化辛卜生公式来说，同法可推得

$$I\approx S_{2M}+\frac{1}{15}(S_{2M}-S_M) \tag{6.48}$$

其中 $S_M$ 为复化辛卜生公式在 $M$ 等分时的积分值，$S_{2M}$为复化辛卜生公式在 $2M$ 等分时的积分值。

对复化柯特斯公式有

$$I\approx C_{2M}+\frac{1}{63}(C_{2M}-C_M) \tag{6.49}$$

其中 $C_M$ 和 $C_{2M}$分别为复化柯特斯公式在 $M$ 等分和 $2M$ 等分时的积分值。

式(6.46)、式(6.48)、式(6.49)亦可改写为

$$I\approx\frac{4}{4-1}T_{2M}-\frac{1}{4-1}T_M \tag{6.50}$$

$$I\approx\frac{4^2}{4^2-1}S_{2M}-\frac{1}{4^2-1}S_M \tag{6.51}$$

$$I\approx\frac{4^3}{4^3-1}C_{2M}-\frac{1}{4^3-1}C_M \tag{6.52}$$

采用上述方法就可以避免事先确定 $M$ 的困难。这种直接用计算结果估计误差的方法称为事后估计误差法。

### 1.4　龙贝格法

由 1.3 看出，把积分区间逐次分半，由式(6.50)、式(6.51)、式(6.52)可获得较精确的积分值。按照事后误差估计式(6.46)知，积分近似值 $T_{2M}$的误差大约等于$(T_{2M}-T_M)/3$。因此，如果用这个误差值作为 $T_{2M}$的一种补偿，就可得到精度更高的改进值。当 $M=1$ 时，按式(6.50)计算得

$$\frac{4}{4-1}T_2-\frac{1}{4-1}T_1=\frac{4}{3}T_2-\frac{1}{3}T_1$$
$$=\frac{b-a}{2}\left[\frac{1}{3}f(a)+\frac{4}{3}f(a+\frac{b-a}{2})+\frac{1}{3}f(b)\right]$$

这正是$[a,b]$区间上的辛卜生求积公式，即

$$S_1=\frac{4}{4-1}T_2-\frac{1}{4-1}T_1 \tag{6.53}$$

同样把 $T_4$ 与 $T_2$ 按式(6.50)可组合成两个辛卜生公式，其相加结果即是复化辛卜生公式 $S_2$

$$S_2=\frac{4}{4-1}T_4-\frac{1}{4-1}T_2 \tag{6.54}$$

依次类推，可得

$$S_{2^k}=\frac{4}{4-1}T_{2^{k+1}}-\frac{1}{4-1}T_{2^k} \tag{6.55}$$

它是 $m=2^k$ 时的复化辛卜生公式。以上说明了由复化梯形公式(6.45)所构成的梯形序列 $T_1$，$T_2$，$T_4$…的线性组合可以形成精度较高的辛卜生序列 $S_1,S_2,S_4,\cdots$。

那么由辛卜生序列 $S_1,S_2,S_4,\cdots$按式(6.51)依次组合是否又能得到精度更高的求积公式呢？取 $M=1$，按式(6.51)有

$$\frac{4^2}{4^2-1}S_2-\frac{1}{4^2-1}S_1=\frac{16}{15}S_2-\frac{1}{15}S_1$$
$$=(b-a)\left[\frac{7}{90}f(a)+\frac{32}{90}f(a+\frac{b-a}{4})+\frac{12}{90}f(a+\frac{b-a}{4}\times 2)+\frac{32}{90}f(a+\frac{b-a}{4}\times 3)+\frac{7}{90}f(b)\right]$$

这正是$[a,b]$区间上的柯特斯公式，即

$$C_1=\frac{4^2}{4^2-1}S_2-\frac{1}{4^2-1}S_1 \tag{6.56}$$

如此类推，可由辛卜生序列按式(6.51)组合成柯特斯序列如下

$$C_{2^k}=\frac{4^2}{4^2-1}S_{2^{k+1}}-\frac{1}{4^2-1}S_{2^k}\qquad(k=0,1,2,\cdots) \tag{6.57}$$

在 $C_{2^k}$已算出的基础上，按式(6.52)继续进行组合为

$$R_{2^k}=\frac{4^3}{4^3-1}C_{2^{k+1}}-\frac{1}{4^3-1}C_{2^k} \tag{6.58}$$

这时 $R_{2^k}$已不再是复化牛顿-柯特斯型的求积公式了，称 $R_{2^k}$ $(k=0,1,2,\cdots)$为龙贝格序列。

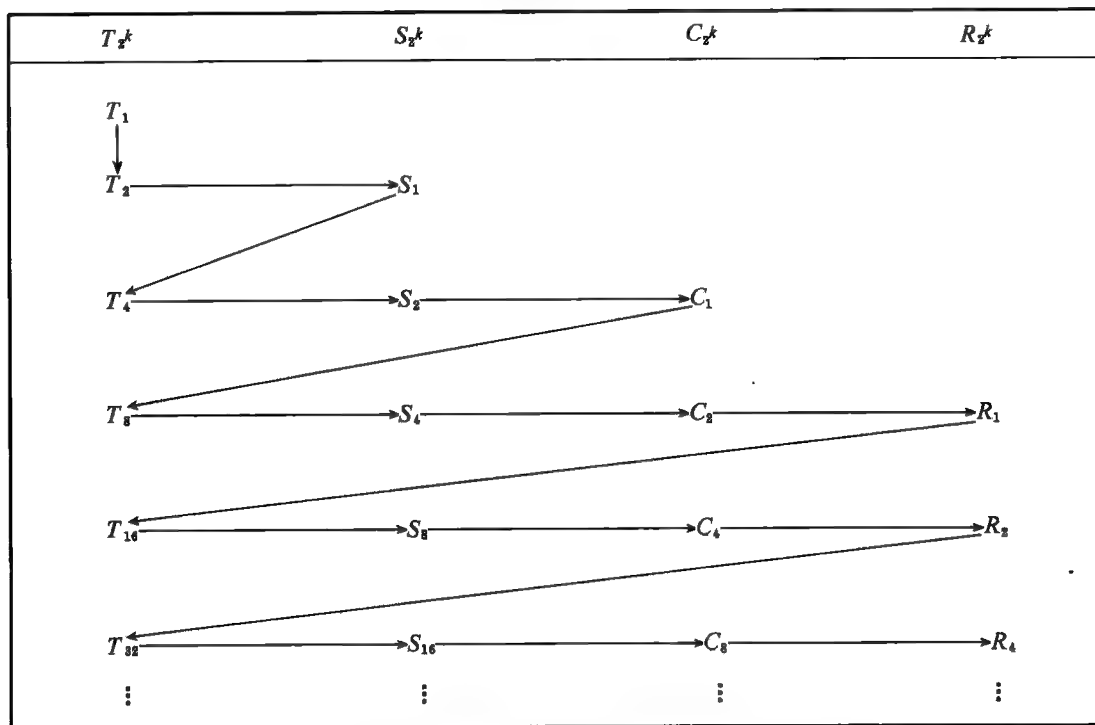
如果按上法继续构造新的序列，其组合中的系数采用 $4^m/(4^m-1)$，$1/(4^m-1)$，则当 $m\geqslant 4$ 时

$$\frac{1}{4^m-1}\leqslant\frac{1}{255}=0.004,\quad \frac{4^m}{4^m-1}\approx 1$$

其组合所得的公式与原来的公式差别不大，故计算时只用到式(6.58)为止，通常称这种求积方法为龙贝格求积方法。也称这个方法为逐次分半加速法或线性加速法。龙贝格法的优点是，系数有规律，无需存储求积系数，且精度较高，计算简易。

实际计算时，并不是做出 $T_{2^k}$ 序列后再做 $S_{2^k}$ 序列，而是在作出 $T_1$，$T_2$ 后计算 $S_1$；在算出 $S_1$、$S_2$ 后计算 $C_1$；在算出 $C_1$、$C_2$ 后计算 $R_1$，以下类推。过程进行到两个相邻结果的绝对误差或相对误差小于所允许的误差限 $\varepsilon$ 时就停止计算。其计算顺序如表 6.4 所示。

表 6.4




**例 6.4** 求 $I=\int_0^1\frac{4}{1+x^2}\mathrm{d}x$ 的近似值，其近似值要求稳定至小数后 5 位。

**解** $f(x)=\frac{4}{1+x^2}$，$[a,b]=[0,1]$，按表 6.4 次序计算如下

$$T_1=\frac{1}{2}[f(0)+f(1)]=\frac{1}{2}[4+2]=3$$

$$T_2=\frac{1}{2}\left[T_1+f\left(\frac{1}{2}\right)\right]=\frac{1}{2}\left[3+\frac{16}{5}\right]=3.1$$

$$S_1=\frac{4}{3}T_2-\frac{1}{3}T_1=\frac{4}{3}\times 3.1-\frac{1}{3}\times 3=3.133\,33$$

$$T_4=\frac{T_2}{2}+\frac{1}{4}\left[f\left(\frac{1}{4}\right)+f\left(\frac{3}{4}\right)\right]=3.131\,18$$

$$S_2=\frac{4T_4-T_2}{3}=3.141\,57$$

$$C_1=\frac{16S_2-S_1}{15}=3.142\ 12$$

$$T_8=\frac{T_4}{2}+\frac{1}{8}\left[f(\frac{1}{8})+f(\frac{3}{8})+f(\frac{5}{8})+f(\frac{7}{8})\right]=3.138\ 99$$

$$S_4=\frac{4T_8-T_4}{3}=3.141\ 59$$

$$C_4=\frac{16S_4-S_2}{15}=3.141\ 59$$

已稳定到小数后第五位，可取 $I=3.141\ 59$。实际上

$$I=\int_0^1\frac{4}{1+x^2}\mathrm{d}x=4\ \arctan x\Big|_0^1=\pi=3.141\ 592\ 6\cdots$$

## 1.5　用差分表达的求积公式

对于由图 5.6 所获得的各种插值公式由 $x'\to x''$ 求其定积分

$$\begin{cases}\displaystyle\int_{x'}^{x''}P_n(x)\mathrm{d}x=h\int_{x'}^{x''}P_n(x_0+th)\mathrm{d}t\\ t=\dfrac{x-x_0}{h}\end{cases}\tag{6.59}$$

上式中对 $P_n(x_0+th)$ 求定积分实际上就是对 $P_n(x_0+th)$ 中所含有的系数 $C_{i+j}^{i}$ 由 $t'\to t''$ 进行积分

$$C_{ij}=\int_{t'}^{t''}C_{i+j}^{i}\mathrm{d}t=\int_{t'}^{t''}\frac{(t+j)(t+j-1)\cdots(t+j-i+1)}{i!}\mathrm{d}t\tag{6.60}$$

显见，只要将插值公式的菱形图 5.6 中的 $C_{i+j}^{i}$ 均换以 $C_{ij}$，就可建立积分域为 $t'\to t''$ 的积分菱形图，再套用图 5.6 相同的使用规则，就能获得以差分表达的 $t'\to t''$ 的求积公式。图 6.1 和图 6.2 就是 $t_0\to t_1$、$t_{-1}\to t_1$ 的积分菱形图，其中 $\Delta_s^m=\Delta^m y_s$。由式(6.59)知，按图 6.1 和图 6.2 所建立的求积公式要用 $h$ 乘之。根据需要，还可以建立其他域上的积分菱形图。

**例 6.5**　建立用差分表示的求积公式 $Q_{01}(0)$。

**解**　$Q_{01}(0)$ 的积分域为 $x_0\to x_1$，使用的插值公式为 $P_0(x)$。由图 6.1 查出相应于 $y_0$ 的系数为 1，因此得以下求积公式

$$\int_{x_0}^{x_1}f(x)\mathrm{d}x\approx Q_{01}(0)=hf(x_0)\tag{6.61}$$

这是矩形公式。

**例 6.6**　建立 $Q_{21}(0)$。

**解**　$Q_{21}(0)$ 的积分域为 $x_0\to x_1$，使用的插值公式为 $P_2(x)$，因此在公式中应取至二阶差分为止。由图 6.1 查出相应于 $y_0$、$\Delta y_0$、$\Delta^2 y_0$ 的系数为 $1$、$\frac{1}{2}$、$-\frac{1}{12}$，因此得以下求积公式

$$\int_{x_0}^{x_1}f(x)\mathrm{d}x\approx Q_{21}(0)=h\left(y_0+\frac{1}{2}\Delta y_0-\frac{1}{12}\Delta^2 y_0\right)\tag{6.62}$$

**例 6.7**　建立与贝塞尔插值公式相应的求积公式。

**解**　由式(5.60)知，贝塞尔插值公式为

$$P_n(x)=\frac{y_0+y_1}{2}+\frac{1}{1!}\left(t-\frac{1}{2}\right)\Delta y_0+\frac{t^{[2]}}{2!}\cdot\frac{\Delta^2 y_{-1}+\Delta^2 y_0}{2}+$$

$$\frac{1}{3!}\left(t-\frac{1}{2}\right)t^{[2]}\Delta^3 y_{-1}+\frac{(t+1)^{[4]}}{4!}\cdot\frac{\Delta^4 y_{-2}+\Delta^4 y_{-1}}{2}+\cdots+$$
$$\frac{1}{(2k-1)!}\left(t-\frac{1}{2}\right)(t+k-2)^{[2k-2]}\Delta^{2k-1}y_{-k+1}+$$
$$\frac{(t+k-1)^{[2k]}}{(2k)!}\cdot\frac{\Delta^{2k}y_{-k}+\Delta^{2k}y_{-k+1}}{2}$$

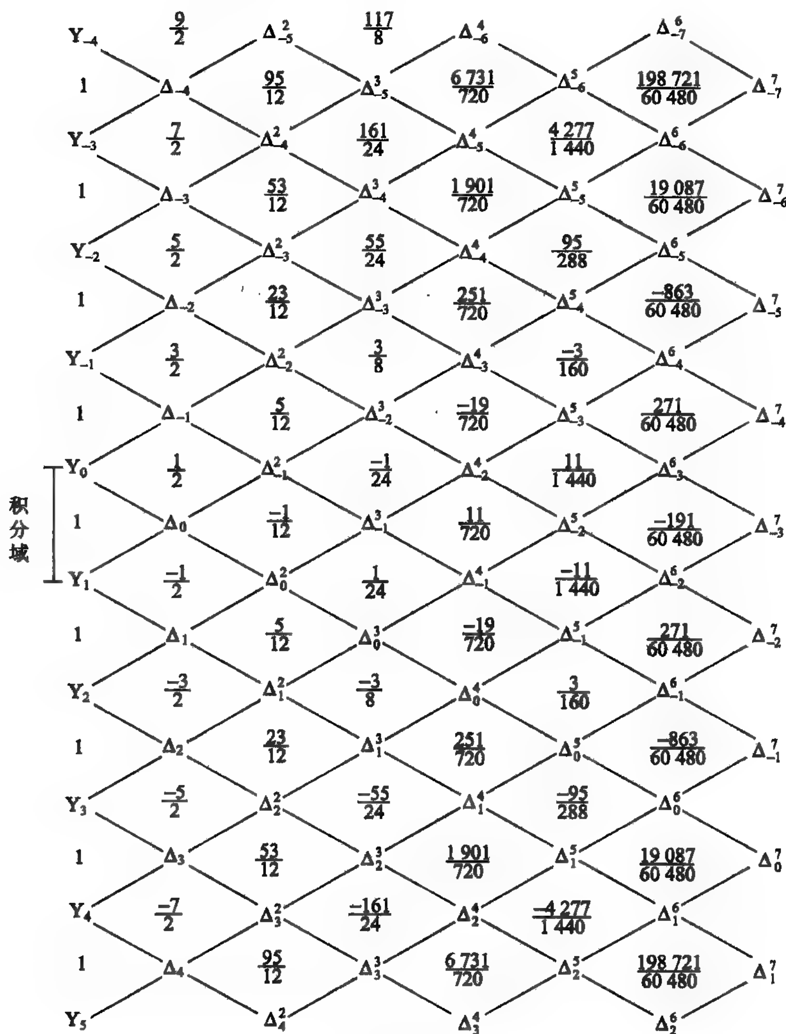



图 6.1

按图 6.1 可写出相应的求积公式

$$\int_{x_0}^{x_0+h} f(x)\mathrm{d}x = h\left[\frac{y_0+y_1}{2}-\frac{1}{12}\cdot\frac{\Delta^2 y_{-1}+\Delta^2 y_0}{2}+\frac{11}{720}\cdot\frac{\Delta^4 y_{-2}+\Delta^4 y_{-1}}{2}-\frac{191}{60\,480}\cdot\frac{\Delta^6 y_{-3}+\Delta^6 y_{-2}}{2}+\cdots+C_k\frac{\Delta^{2k}y_{-k}+\Delta^{2k}y_{-k+1}}{2}\right]+R \quad (6.63)$$

其中

$$C_k = \frac{1}{(2k)!}\int_0^1 (t+k-1)\cdots(t-k)\mathrm{d}t$$

$$R = \int_{x_0}^{x_0+h} R_{2k}(x)\mathrm{d}x$$

**例 6.8**　求 $\int_{x_{-1}}^{x_{-1}+2h} y'(x)\mathrm{d}x$ 的计算公式。

**解**　图 6.2 为 $y'$ 的积分菱形图，若插值多项式取为二次，则插值多项式中差分的最高阶数为 2。由图 6.2 中查出相应于 $y'_{-1}$、$\Delta y'_{-1}$、$\Delta^2 y'_{-1}$ 的系数为 2、2、$\frac{1}{3}$，故得以下求积公式

$$\int_{x_{-1}}^{x_1} y'(x)\mathrm{d}x = h[2y'_{-1} + 2\Delta y'_{-1} + \frac{1}{3}\Delta^2 y'_{-1}]$$

所以

$$y_1 = y_{-1} + h[2y'_{-1} + 2\Delta y'_{-1} + \frac{1}{3}\Delta^2 y'_{-1}] \tag{6.64}$$

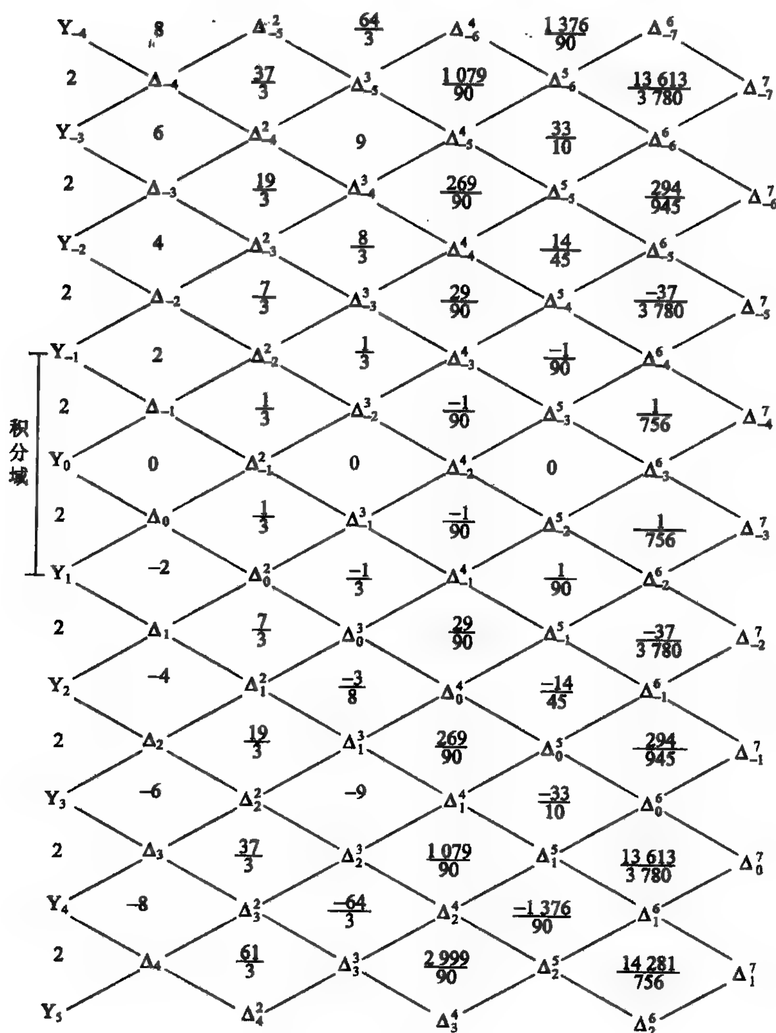


图 6.2

## 1.6 切比雪夫求积公式

**定理** 设代数方程

$$f(t) = t^n + a_1 t^{n-1} + \cdots + a_{n-1}t + a_n = 0 \tag{6.65}$$

的根为 $t_1, t_2, \cdots, t_n$，记

$$S_r = t_1^r + t_2^r + \cdots + t_n^r = \sum_{i=1}^{n} t_i^r \qquad (r = 1, 2, \cdots) \tag{6.66}$$

则有如下关系式成立

$$\begin{cases} S_1 + a_1 = 0 \\ S_2 + a_1 S_1 + 2a_2 = 0 \\ S_3 + a_1 S_2 + a_2 S_1 + 3a_3 = 0 \\ \cdots \\ S_n + a_1 S_{n-1} + a_2 S_{n-2} + \cdots + n a_n = 0 \end{cases} \tag{6.67}$$

或写成 $$S_m + a_1 S_{m-1} + a_2 S_{m-2} + \cdots + a_{m-1} S_1 + m a_m = 0 \qquad (m = 1, 2, \cdots, n) \tag{6.68}$$

**证** 对下式

$$f(t) = (t - t_1)(t - t_2)\cdots(t - t_n)$$

求导得

$$\begin{aligned} f'(t) &= \sum_{i=1}^{n} (t - t_1)\cdots(t - t_{i-1})(t - t_{i+1})\cdots(t - t_n) \\ &= \sum_{i=1}^{n} \frac{(t - t_1)\cdots(t - t_{i-1})(t - t_i)(t - t_{i+1})\cdots(t - t_n)}{(t - t_i)} \\ &= \sum_{i=1}^{n} \frac{f(t)}{t - t_i} \end{aligned} \tag{6.69}$$

由 $f(t_i) = 0$，故知

$$\begin{aligned} \frac{f(t)}{t - t_i} &= \frac{f(t) - f(t_i)}{t - t_i} \\ &= \frac{t^n - t_i^n}{t - t_i} + a_1 \frac{t^{n-1} - t_i^{n-1}}{t - t_i} + \cdots + a_{n-1} \frac{t - t_i}{t - t_i} \\ &= t^{n-1} + t_i t^{n-2} + t_i^2 t^{n-3} + t_i^3 t^{n-4} + \cdots + t_i^{n-1} + a_1 (t^{n-2} + t_i t^{n-3} + t_i^2 t^{n-4} + \cdots + t_i^{n-2}) + \\ &\quad a_2 (t^{n-3} + t_i t^{n-4} + \cdots + t_i^{n-3}) + a_3 (t^{n-4} + \cdots + t_i^{n-4}) + \cdots + a_{n-1} \\ &= t^{n-1} + (t_i + a_1) t^{n-2} + (t_i^2 + a_1 t_i + a_2) t^{n-3} + (t_i^3 + a_1 t_i^2 + a_2 t_i + a_3) t^{n-4} + \cdots + \\ &\quad (t_i^{n-1} + a_1 t_i^{n-2} + a_2 t_i^{n-3} + \cdots + a_{n-1}) \end{aligned} \tag{6.70}$$

代入式(6.69) 后得

$$\begin{aligned} f'(t) = &\, n t^{n-1} + (S_1 + n a_1) t^{n-2} + (S_2 + a_1 S_1 + n a_2) t^{n-3} + \cdots + \\ &(S_{n-1} + a_1 S_{n-2} + \cdots + a_{n-2} S_1 + n a_{n-1}) \end{aligned} \tag{6.71}$$

对式(6.65) 直接求导得

$$f'(t) = n t^{n-1} + (n-1) a_1 t^{n-2} + (n-2) a_2 t^{n-3} + \cdots + a_{n-1} \tag{6.72}$$

比较式(6.71) 和式(6.72) 得

$$\begin{cases} S_1 + na_1 = (n-1)a_1 \\ S_2 + a_1 S_1 + na_2 = (n-2)a_2 \\ \cdots \\ S_{n-1} + a_1 S_{n-2} + \cdots + a_{n-2} S_1 + na_{n-1} = a_{n-1} \end{cases}$$

即得

$$\begin{cases} S_1 + a_1 = 0 \\ S_2 + a_1 S_1 + 2a_2 = 0 \\ \cdots \\ S_{n-1} + a_1 S_{n-2} + \cdots + a_{n-2} S_1 + (n-1)a_{n-1} = 0 \end{cases} \tag{6.73}$$

再由式(6.65)对 $t_i(i=1,2,\cdots,n)$求和得

$$\sum_{i=1}^{n} f(t_i) = \sum_{i=1}^{n} t_i^n + a_1 \sum_{i=1}^{n} t_i^{n-1} + \cdots + \sum_{i=1}^{n} a_n$$
$$= S_n + a_1 S_{n-1} + \cdots + na_n = 0 \tag{6.74}$$

综合式(6.73)与式(6.74),即知式(6.68)成立。

当已知 $a_1,a_2,\cdots,a_n$ 的情况下,利用(6.68)式可逐次推算出 $S_1,S_2,\cdots,S_n$ 的数值,反之亦成立。

以下讨论求积公式

$$\int_{-1}^{+1} f(t)\mathrm{d}t \approx \sum_{i=1}^{n} c_i f(t_i) \tag{6.75}$$

切比雪夫提出:①取相同的系数值 $c_1=c_2=\cdots=c_n=c$;②要求式(6.75)对任意不大于 $n$ 次的多项式精确成立。

按条件①可将式(6.75)写成

$$\int_{-1}^{+1} f(t)\mathrm{d}t \approx c\sum_{i=1}^{n} f(t_i) \tag{6.76}$$

假定任意 $n$ 次多项式为

$$P_n(t) = b_0 + b_1 t + \cdots + b_n t^n \tag{6.77}$$

根据条件②,如将式(6.77)代入式(6.76)后两边应恒等,即

$$\int_{-1}^{+1} (b_0 + b_1 t + \cdots + b_n t^n)\mathrm{d}t = c\sum_{i=1}^{n} (b_0 + b_1 t_i + \cdots + b_n t_i^n)$$

由此得到

$$2\left(b_0 + \frac{b_2}{3} + \frac{b_4}{5} + \frac{b_6}{7} + \cdots\right) = c\left(nb_0 + b_1 \sum_{i=1}^{n} t_i + b_2 \sum_{i=1}^{n} t_i^2 + \cdots + b_n \sum_{i=1}^{n} t_i^n\right)$$

比较左右两边 $b_i$ 的系数得以下关系式

$$\begin{cases} c = \dfrac{2}{n} \\ S_1 = t_1 + t_2 + \cdots + t_n = 0 \\ S_2 = t_1^2 + t_2^2 + \cdots + t_n^2 = \dfrac{n}{3} \\ S_3 = t_1^3 + t_2^3 + \cdots + t_n^3 = 0 \\ S_4 = t_1^4 + t_2^4 + \cdots + t_n^4 = \dfrac{n}{5} \\ \cdots \\ S_n = t_1^n + t_2^n + \cdots + t_n^n = \dfrac{n}{2}\,\dfrac{1-(-1)^{n+1}}{n+1} \end{cases} \tag{6.78}$$

将上述 $S_i(i=1,2,\cdots,n)$ 代入式(6.67)得

$$\begin{cases} a_1=0 \\ \dfrac{n}{3}+2a_2=0 \\ a_3=0 \\ \dfrac{n}{5}+\dfrac{n}{3}a_2+4a_4=0 \\ a_5=0 \\ \dfrac{n}{7}+\dfrac{n}{5}a_2+\dfrac{n}{3}a_4+6a_6=0 \\ a_7=0 \\ \cdots \end{cases} \tag{6.79}$$

令 $n=1,2,\cdots,7,9$，由式(6.79)分别解得八组解 $a_i(i=1,2,\cdots,n)$ 各值，代入式(6.65)得到以下相应的八个代数方程

$$\begin{cases} n=1: & t=0 \\ n=2: & t^2-\dfrac{1}{3}=0 \\ n=3: & t^3-\dfrac{1}{2}t=0 \\ n=4: & t^4-\dfrac{2}{3}t^2+\dfrac{1}{45}=0 \\ n=5: & t^5-\dfrac{5}{6}t^3+\dfrac{7}{12}t=0 \\ n=6: & t^6-t^4+\dfrac{1}{5}t^2-\dfrac{1}{105}=0 \\ n=7: & t^7-\dfrac{7}{6}t^5+\dfrac{119}{360}t^3-\dfrac{149}{1\,480}=0 \\ n=9: & t^9-\dfrac{3}{2}t^7+\dfrac{27}{40}t^5-\dfrac{57}{560}t^3+\dfrac{53}{22\,400}t=0 \end{cases} \tag{6.80}$$

以上各个方程的解列入表 6.5 中以备查用。当 $n=8$ 时无实解，$n\geqslant 10$ 时不能解●。

**表 6.5**

| $n$ | $t_i$ | $R_n$ |
|---|---|---|
| 1 | $t_1=0$ | $R_1=\dfrac{1}{3}f''(\xi)$ |
| 2 | $t_2=-t_1=0.577\,350$ | $R_2=\dfrac{1}{135}f^{(4)}(\xi)$ |
| 3 | $t_3=-t_1=0.707\,107$，$t_2=0$ | $R_3=\dfrac{1}{360}f^{(4)}(\xi)$ |
| 4 | $t_4=-t_1=0.794\,654$，$t_3=-t_2=0.187\,592$ | $R_4=\dfrac{2}{425\,25}f^{(6)}(\xi)$ |

● 参看 N·Л·纳唐松著《函数构造论》(下册)147 页和 153 页，科学出版社 1959 年出版。

续表

| $n$ | $t_i$ | $R_n$ |
|---|---|---|
| 5 | $t_5=-t_1=0.832\,498$，$t_3=0$，$t_4=-t_2=0.374\,541$ | $R_5=\frac{13}{544\,320}f^{(6)}(\xi)$ |
| 6 | $t_6=-t_1=0.866\,247$，$t_5=-t_2=0.422\,519$，$t_4=-t_3=0.266\,635$ | $R_6=\frac{1}{3\,969\,000}f^{(8)}(\xi)$ |
| 7 | $t_7=-t_1=0.883\,862$，$t_6=-t_2=0.529\,657$，$t_5=-t_3=0.323\,912$，$t_4=0$ | $R_7=\frac{281}{1\,959\,552\,000}f^{(8)}(\xi)$ |
| 9 | $t_9=-t_1=0.911\,589$，$t_8=-t_2=0.601\,019$，$t_7=-t_3=0.528\,762$，$t_6=-t_4=0.167\,906$，$t_5=0$ | $R_9=\frac{747\,49}{11\,200\times9\times11!}f^{(10)}(\xi)$ |

切比雪夫求积公式的余式为

$$R_n=\int_{-1}^{+1}f[x,\ 0,\ x_1,\ -x_1,\ \cdots,\ x_m,\ -x_m]\,x\prod_{k=1}^{m}(x^2-x_k^2)\mathrm{d}x \tag{6.81}$$

其中当 $n$ 为偶数时，$m=\frac{n}{2}$；$n$ 为奇数时，$m=\frac{n-1}{2}$。若 $f(x)$ 具有 $2m+1$ 阶导数，则

$$R_n=\frac{1}{(2m+1)!}\int_{-1}^{+1}f^{(2m+1)}(\xi)x\prod_{k=1}^{m}(x^2-x_k^2)\mathrm{d}x \tag{6.82}$$

当积分区间为$[a,\ b]$时，可令

$$x=\frac{a+b}{2}+\frac{b-a}{2}t \tag{6.83}$$

则

$$\int_a^b f(x)\mathrm{d}x=\int_{-1}^{+1}\frac{b-a}{2}f\left(\frac{a+b}{2}+\frac{b-a}{2}t\right)\mathrm{d}t$$

$$=\int_{-1}^{+1}F(t)\mathrm{d}t\approx\frac{2}{n}\sum_{i=1}^{n}F(t_i) \tag{6.84}$$

其中

$$F(t)=\frac{b-a}{2}f\left(\frac{a+b}{2}+\frac{b-a}{2}t\right) \tag{6.85}$$

一个求积公式如果对任意 $n$ 次多项式精确成立(这时余式恒为 0)而对任何大于 $n$ 次的多项式不精确成立，就称该求积公式具有 $n$ 次代数精确度。一般说来，当 $f(x)$ 不是次数小于 $n$ 的多项式时，该求积公式就不一定精确了。求积公式的准确度的一种度量方法是考察其公式的截断误差。此外，一个求积公式的代数精确度也是该求积公式近似程度的一种度量，因为当我们用插值多项式 $P_n(x)$ 近似 $f(x)$ 计算积分近似值时，如果求积公式的代数精确度 $n$ 愈大，则截断误差 $P_n(x)-f(x)$ 愈小，因此相应的积分近似值愈精确。由于 $P_n(t)$ 的 $n$ 次拉格朗日插值公式 $L_n(t)\equiv P_n(t)$，故牛顿-柯特斯求积公式至少具有 $n$ 次代数精确度。根据切比雪夫求积公式的建立方法可知，切比雪夫求积公式具有 $n$ 次代数精确度。由求积公式的余式可见，若余式中函数的导数阶数越高，其代数精确度也越高。

**例 6.9** 按切比雪夫求积公式($n=4$)计算下述积分近似值

$$I=\int_0^1\frac{\sin x}{x}\mathrm{d}x$$

**解** 按式(6.83)得

$$x=\frac{1}{2}(1+t)$$

由表6.5查得 $n=4$ 时 $t_i$ 为

$$t_1=-0.794\ 654,\ t_2=-0.187\ 592,\ t_3=0.187\ 592,\ t_4=0.794\ 65$$

相应的 $x_i$，$f(x_i)$为

$$x_1=\frac{1}{2}(1+t_1)=0.102\ 673,\quad f(x_1)=\frac{\sin x_1}{x_1}=0.998\ 244$$

$$x_2=\frac{1}{2}(1+t_2)=0.406\ 204,\quad f(x_2)=\frac{\sin x_2}{x_2}=0.972\ 726$$

$$x_3=\frac{1}{2}(1+t_3)=0.593\ 796,\quad f(x_3)=\frac{\sin x_3}{x_3}=0.942\ 262$$

$$x_4=\frac{1}{2}(1+t_4)=0.897\ 327,\quad f(x_4)=\frac{\sin x_4}{x_4}=0.871\ 101$$

按式(6.84)求取 $I$ 的近似值

$$\begin{aligned}I&\approx\frac{1}{4}[f(x_1)+f(x_2)+f(x_3)+f(x_4)]\\&=\frac{1}{4}[0.998\ 244+0.972\ 726+0.942\ 262+0.871\ 101]\\&=\frac{1}{4}\times 3.784\ 333=0.946\ 083\end{aligned}$$

下面估计截断误差的大小。由于

$$f(x)=\frac{\sin x}{x}=\int_0^1\cos tx\,dt$$

所以

$$f^{(6)}(x)=\frac{d^6}{dx^6}\left(\frac{\sin x}{x}\right)=\int_0^1\frac{d^6}{dx^6}(\cos tx)dt=-\int_0^1 t^6\cos tx\,dt$$

于是有

$$\left|F^{(6)}(x)\right|=\frac{1}{2}\left|f^{(6)}(x)\right|\leqslant\frac{1}{2}\int_0^1\max_{0\leqslant x\leqslant 1}\left|t^6\cos tx\right|dt\leqslant\frac{1}{2}\int_0^1 t^6 dt=\frac{1}{14}$$

因此

$$\left|R_4\right|=\frac{2}{42\ 525}\left|F^{(6)}(\xi)\right|\leqslant\frac{2}{42\ 525}\times\frac{1}{14}\approx 0.3\times 10^{-5}$$

绝对误差限 $\Delta f_i$ 可估计如下

$$\Delta f_i=\left[\left|\frac{0.5\times 10^{-6}}{\sin x_i}\right|+\left|\frac{0.5\times 10^{-6}}{x_i}\right|\right]\cdot\left|f(x_i)\right|=\begin{cases}0.97\times 10^{-5},\ i=1\\0.24\times 10^{-5},\ i=2\\0.10\times 10^{-5},\ i=3\\0.10\times 10^{-5},\ i=4\end{cases}$$

所以结果的舍入误差为

$$|\in|\leqslant\frac{1}{4}(0.97\times 10^{-5}+0.24\times 10^{-5}+0.10\times 10^{-5}+0.10\times 10^{-5})=0.35\times 10^{-5}$$

总误差为 $$|\varepsilon|\leqslant 0.3\times 10^{-5}+0.35\times 10^{-5}=0.65\times 10^{-5}<0.5\times 10^{-4}$$

取 $I\approx 0.946\ 1$。

## 1.7　高斯求积公式

高斯求积公式求积分近似值的方法在于不固定系数，也不固定节点 $x_i$，而是把节点视为 $n$ 个可以自由选择的参数，选择这些参数的原则是使求积公式

$$\int_{-1}^{1} f(t)\mathrm{d}t = \sum_{i=1}^{n} c_i f(t_i) \tag{6.86}$$

对次数尽可能高的多项式精确成立。

在[−1,1]区间上，先假定插值节点为 $t_1, t_2, \cdots, t_n$，相应的函数值为 $y_1, y_2, \cdots, y_n$，建立拉格朗日插值公式

$$L_{n-1}(t) = \sum_{i=1}^{n} \frac{\prod_n(t)}{\prod'_n(t_i)(t-t_i)} y_i \tag{6.87}$$

其中 $\Pi_n(t)=(t-t_1)(t-t_2)\cdots(t-t_n)$。用 $L_{n-1}(t)$ 近似 $f(t)$，在[−1,1]区间上计算 $f(t)$ 的积分近似值

$$\begin{aligned}\int_{-1}^{1} f(t)\mathrm{d}t &\approx \int_{-1}^{1} L_{n-1}(t)\mathrm{d}t \\ &= \int_{-1}^{1} \sum_{i=1}^{n} \frac{\prod_n(t)}{\prod_n'(t_i)(t-t_i)} y_i \mathrm{d}t \\ &= \sum_{i=1}^{n} \left[\int_{-1}^{1} \frac{\prod_n(t)}{\prod_n'(t_i)(t-t_i)} \mathrm{d}t\right] y_i \\ &= \sum_{i=1}^{n} c_i y_i \end{aligned} \tag{6.88}$$

其中

$$c_i = \int_{-1}^{1} \frac{\prod_n(t)}{\prod'_n(t_i)(t-t_i)} \mathrm{d}t \quad (i=1,\ 2,\ \cdots,\ n) \tag{6.89}$$

为了提高精度，再引入 $m$ 个新的节点，设为 $t_{n+1}, t_{n+2}, \cdots, t_{n+m}$，对应于节点 $t_i(i=1, 2, \cdots, n, n+1, \cdots, n+m)$ 的拉格朗日插值公式为

$$L_{n+m-1}(t) = \sum_{i=1}^{n} a_i^{(1)}(t) y_i + \sum_{i=n+1}^{n+m} a_i^{(2)}(t) y_i \tag{6.90}$$

其中

$$\begin{aligned} a_i^{(1)}(t) &= \frac{\prod_n(t)}{\prod'_n(t_i)(t-t_i)} \cdot \frac{(t-t_{n+1})(t-t_{n+2})\cdots(t-t_{n+m})}{(t_i-t_{n+1})(t_i-t_{n+2})\cdots(t_i-t_{n+m})} \\ &= \frac{\prod_n(t)}{\prod'_n(t_i)(t-t_i)} \left[\frac{t-t_{n+1}}{t_i-t_{n+1}} \frac{t-t_{n+2}}{t_i-t_{n+2}} \cdots \frac{t-t_{n+m}}{t_i-t_{n+m}}\right] \\ &= \frac{\prod_n(t)}{\prod'_n(t_i)(t-t_i)} \left[\left(1+\frac{t-t_i}{t_i-t_{n+1}}\right)\left(1+\frac{t-t_i}{t_i-t_{n+2}}\right)\cdots\left(1+\frac{t-t_i}{t_i-t_{n+m}}\right)\right] \\ &= \frac{\prod_n(t)}{\prod'_n(t_i)(t-t_i)} \left[1+k_0(t-t_i)+k_1(t-t_i)^2+\cdots+k_{m-1}(t-t_i)^m\right] \end{aligned}$$

$$= \frac{\prod_n(t)}{\prod_n'(t_i)(t-t_i)} + \frac{\prod_n(t)}{\prod'_n(t_i)}[k_0 + k_1(t-t_i) + \cdots + k_{m-1}(t-t_i)^{m-1}]$$

$$= \frac{\prod_n(t)}{\prod'_n(t_i)(t-t_i)} + \prod_n(t) \cdot Q_{m-1}(t) \qquad (6.91)$$

式中 $$Q_{m-1}(t) = \frac{1}{\prod'_n(t_i)}[k_0 + k_1(t-t_i) + \cdots + k_{m-1}(t-t_i)^{m-1}]$$

为 $t$ 的 $m-1$ 次多项式。而

$$a_i^{(2)}(t) = \frac{\prod_n(t)(t-t_{n+1})\cdots(t-t_{i-1})(t-t_{i+1})\cdots(t-t_{n+m})}{\prod_n(t_i)(t_i-t_{n+1})\cdots(t_i-t_{i-1})(t_i-t_{i+1})\cdots(t_i-t_{n+m})}$$

$$= \prod_n(t) \cdot \tilde{Q}_{m-1}(t) \qquad (6.92)$$

式中 $$\tilde{Q}_{m-1}(t) = \frac{(t-t_{n+1})\cdots(t-t_{i-1})(t-t_{i+1})\cdots(t-t_{n+m})}{\prod_n(t_i)(t_i-t_{n+1})\cdots(t_i-t_{i-1})(t_i-t_{i+1})\cdots(t_i-t_{n+m})}$$

亦是 $t$ 的 $m-1$ 次多项式。今以 $L_{n+m-1}(t)$ 近似 $f(t)$ 在$[-1,1]$区间上积分值得

$$\int_{-1}^{1} f(t)\mathrm{d}t \approx \int_{-1}^{1} L_{n+m-1}(t)\mathrm{d}t = \sum_{i=1}^{n}\left[\int_{-1}^{1}\frac{\prod_n(t)}{\prod'_n(t_i)(t-t_i)}\mathrm{d}t\right]y_i + \sum_{i=1}^{n}\left[\int_{-1}^{1}\prod_n(t)Q_{m-1}(t)\mathrm{d}t\right]y_i +$$

$$\sum_{i=n+1}^{n+m}\left[\int_{-1}^{1}\prod_n(t)\tilde{Q}_{m-1}(t)\mathrm{d}t\right]y_i \qquad (6.93)$$

求积公式(6.93)虽然比求积公式(6.88)有较高的代数精确度,但却增添了如下的两项运算量

$$A_1 = \sum_{i=1}^{n}\left[\int_{-1}^{1}\prod_n(t)Q_{m-1}(t)\mathrm{d}t\right]y_i$$

$$A_2 = \sum_{i=n+1}^{n+m}\left[\int_{-1}^{1}\prod_n(t)\tilde{Q}_{m-1}(t)\mathrm{d}t\right]y_i \qquad (6.94)$$

是否能选取适当的多项式 $P_n(t)$ 作为 $\prod_n(t)$,使式(6.94)的 $A_1=0, A_2=0$ 同时又使式(6.93)具有尽可能高的代数精确度呢?这个问题的答案是取 $n$ 次勒让德多项式

$$P_n(t) = \frac{1}{2^n n!}\frac{\mathrm{d}^n}{\mathrm{d}x^n}(x^2-1)^n \qquad (6.95)$$

作为 $\prod_n(t)$ 及取 $m=n$ 时就可以达到上述目的。这是因为勒让德多项式具有以下性质。

① 勒让德多项式是$[-1,1]$区间上的正交函数组,即

$$\int_{-1}^{1} P_n(t)P_m(t)\mathrm{d}t \begin{cases} =0, & m \neq n \\ \neq 0, & m=n \text{ 且 } m,\ n=0,\ 1,\ 2,\ \cdots \end{cases} \qquad (6.96)$$

② 对一切 $k<n$,有

$$\int_{-1}^{1} P_n(t)Q_k(t)\mathrm{d}t = 0 \qquad (6.97)$$

式中,$Q_k(t)$ 是 $t$ 的任意 $k$ 次多项式。这是由于 $Q_k(t)$ 可用勒让德多项式表示为如下的线性组合

$$Q_k(t) = c_0P_k(t) + c_1P_{k-1}(t) + \cdots + c_kP_0(t) \qquad (6.98)$$

则 $$\int_{-1}^{1} P_n(t)Q_k(t)\mathrm{d}t = \int_{-1}^{1} P_n(t)\left[\sum_{i=0}^{k}c_iP_{k-i}(t)\right]\mathrm{d}t$$

$$= \sum_{i=0}^{k} c_i \int_{-1}^{1} P_n(t) P_{k-i}(t) \mathrm{d}t = 0$$

③ $n$ 次勒让德多项式在$[-1,1]$内具有 $n$ 个不同的零点(见表 6.6)，利用这些零点可将 $n$ 次勒让德多项式 $P_n(t)$ 表为

$$P_n(t) = (t-t_1)(t-t_2)\cdots(t-t_n) \tag{6.99}$$

根据上述性质，由公式(6.94)可见，若取 $m=n$ 及取 $n$ 次勒让德多项式 $P_n(t)$ 的零点 $t_1, t_2, \cdots, t_n$ 作为插值节点时，就可使 $A_1=A_2=0$，而公式(6.93)化为

$$\int_{-1}^{1} f(t)\mathrm{d}t \approx \int_{-1}^{1} L_{2n-1}(t)\mathrm{d}t = \sum_{i=1}^{n} w_i y_i \tag{6.100}$$

其中

$$w_i = \int_{-1}^{1} \frac{P_n(t)}{P_n{'}(t_i)(t-t_i)}\mathrm{d}t \quad (i=1,2,\cdots,n) \tag{6.101}$$

公式(6.100)称为高斯求积公式，它具有 $2n-1$ 次的代数精确度。其余式为

$$R_n = \frac{f^{(2n)}(\xi)}{(2n)!}\int_{-1}^{1} P_n^2(t)\mathrm{d}t = 2^{2n+1}\frac{(n!)^4}{[(2n!)]^3}\frac{f^{(2n)}(\xi)}{2n+1},\ |\xi|<1 \tag{6.102}$$

与其他求积公式比较，高斯求积公式兼有使用节点数目少且精度高的双重优点。由于达到同样精度时，高斯求积公式所需的节点数目少，相应地参与运算的函数值也少，所以舍入误差也小。关于高斯求积公式中所使用的节点和系数值、余式表达式均列入表 6.6 中以备查用。对于下述积分

$$\int_a^b f(x)\mathrm{d}x$$

可令

$$x = \frac{a+b}{2} + \frac{b-a}{2}t$$

则有

$$\int_a^b f(x)\mathrm{d}x = \frac{b-a}{2}\int_{-1}^{1} f\left(\frac{a+b}{2}+\frac{b-a}{2}t\right)\mathrm{d}t \tag{6.103}$$

表 6.6

| $n$ | $\pm t_i$ | $w_i$ | $R_n$ |
|---|---|---|---|
| 2 | 0.577 350 269 2 | 1 | $R_2=\frac{1}{135}f^{(4)}(\xi)$ |
| 3 | 0<br>0.774 596 692 | 0.888 888 888 9<br>0.555 555 555 6 | $R_3=\frac{1}{157\ 50}f^{(6)}(\xi)$ |
| 4 | 0.339 981 043 5<br>0.861 136 311 6 | 0.652 145 154 9<br>0.347 854 845 1 | $R_4=\frac{1}{347\ 287\ 5}f^{(8)}(\xi)$ |
| 5 | 0<br>0.538 469 310 1<br>0.906 179 845 9 | 0.588 888 888 9<br>0.478 628 670 5<br>0.236 926 885 1 | $R_5=\frac{1}{123\ 773\ 265\ 0}f^{(10)}(\xi)$ |
| 6 | 0.238 619 186 1<br>0.661 209 386 5<br>0.932 469 514 2 | 0.467 913 934 6<br>0.360 761 573 0<br>0.171 324 492 4 | $R_6=\frac{1}{648\ 984\ 486\ 150}f^{(12)}(\xi)$ |
| 7 | 0<br>0.405 845 151 4<br>0.741 531 185 6<br>0.949 107 912 3 | 0.471 959 183 7<br>0.381 830 050 5<br>0.279 705 391 5<br>0.129 484 966 2 | $R_7=\frac{1}{470\ 050\ 192\ 111\ 500}f^{(14)}(\xi)$ |

相应的高斯求积公式为

$$\int_a^b f(x)\mathrm{d}x \approx \frac{b-a}{2}\sum_{i=1}^{n} w_i f(x_i) \tag{6.104}$$

$$x_i = \frac{a+b}{2} + \frac{b-a}{2}t_i \qquad (i=1,2,\cdots,n) \tag{6.105}$$

式中，$t_i$ 为勒让德多项式 $P_n(t)$ 的零点。式(6.104)的余式为

$$R_n = \frac{(b-a)^{2n+1}(n!)^4 f^{(2n)}(\xi)}{[(2n)!]^3(2n+1)} = \left(\frac{b-a}{2}\right)^{2n+1}\frac{2^{2n+1}(n!)^4 f^{(2n)}(\xi)}{[(2n)!]^3(2n+1)},\ \xi\in(a,b) \tag{6.106}$$

按上式可得

$$R_2 = \frac{1}{135}\left(\frac{b-a}{2}\right)^5 f^{(4)}(\xi)$$

$$R_3 = \frac{1}{15\ 750}\left(\frac{b-a}{2}\right)^7 f^{(6)}(\xi)$$

$$R_4 = \frac{1}{3\ 472\ 875}\left(\frac{b-a}{2}\right)^9 f^{(8)}(\xi)$$

$$R_5 = \frac{1}{1\ 237\ 732\ 650}\left(\frac{b-a}{2}\right)^{11} f^{(10)}(\xi) \tag{6.107}$$

等。

**例 6.10** 利用高斯求积公式($n=3$)计算下列积分

$$I = \int_0^1 \sqrt{1+2x}\,\mathrm{d}x$$

的近似值。

**解** 按式(6.105)计算以下节点值及函数值

$$x_1 = \frac{1}{2} + \frac{1}{2}t_1 = \frac{1}{2} + \frac{1}{2}\times(-0.774\ 60) = 0.112\ 70$$

$$x_2 = \frac{1}{2} + \frac{1}{2}t_2 = \frac{1}{2} + \frac{1}{2}\times 0 = 0.500\ 00$$

$$x_3 = \frac{1}{2} + \frac{1}{2}t_3 = \frac{1}{2} + \frac{1}{2}\times 0.774\ 60 = 0.887\ 30$$

$$f(x_1) = \sqrt{1+2x_1} = 1.106\ 98$$

$$f(x_2) = \sqrt{1+2x_2} = 1.414\ 21$$

$$f(x_3) = \sqrt{1+2x_3} = 1.665\ 71$$

得 $$I \approx \frac{1-0}{2}(0.555\ 56\times 1.106\ 98 + 0.888\ 89\times 1.414\ 21 + 0.555\ 56\times 1.665\ 71)$$

$$= 1.398\ 70$$

因 $f(x)=\sqrt{1+2x}=(1+2x)^{\frac{1}{2}}$，所以

$$f^{(6)}(x) = \frac{1}{2}\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\left(-\frac{5}{2}\right)\left(-\frac{7}{2}\right)\left(-\frac{9}{2}\right)2^6(1+2x)^{-\frac{11}{2}}$$

$$= -945(1+2x)^{-\frac{11}{2}}$$

则 $$\max_{0\leqslant x\leqslant 1}\left|f^{(6)}(x)\right| = 945$$

$$R_3 = \frac{1}{157\ 50}\left(\frac{1-0}{2}\right)^7 \left|f^{(6)}(\xi)\right| \leqslant \frac{945}{15\ 750}\left(\frac{1}{2}\right)^7$$

$$\approx \frac{1}{2\ 000} = 0.5 \times 10^{-3}$$

$I$ 的舍入误差限可估计如下

$$|\in| \leqslant \frac{1}{2}[(0.555\ 56 + 1.106\ 98) \times 0.5 \times 10^{-5} + (0.888\ 89 + 1.414\ 21) \times 0.5 \times 10^{-5} + (0.555\ 56 + 1.665\ 71) \times 0.5 \times 10^{-5}] = 0.15 \times 10^{-4}$$

总误差为

$$\varepsilon = 0.5 \times 10^{-3} + 0.15 \times 10^{-4} \approx 0.5 \times 10^{-3}$$

所以可取

$$I \approx 1.399$$

### 1.8　重积分的求积公式

前面所述及的求积公式，也可以应用到多重积分的计算上去。例如求积分

$$\begin{aligned} I &= \int_a^b \int_c^d f(x, y)\mathrm{d}x\mathrm{d}y = \int_c^d \left[\int_a^b f(x, y)\mathrm{d}x\right]\mathrm{d}y \\ &\approx \int_c^d \left[\sum_{i=1}^m c_i f(x_i, y)\right]\mathrm{d}y = \sum_{i=1}^m c_i \int_c^d f(x_i, y)\mathrm{d}y \\ &\approx \sum_{i=1}^m \sum_{j=1}^n c_i c_j f(x_i, y_j) = \sum_{i=1}^m \sum_{j=1}^n \lambda_{ij} f_{ij} \end{aligned} \tag{6.108}$$

仿之，三重积分的求积公式为

$$\int_a^b \int_c^d \int_e^f f(x, y, z)\mathrm{d}x\mathrm{d}y\mathrm{d}z \approx \sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^r \lambda_{ijk} f_{ijk} \tag{6.109}$$

余类推。

例如对于 $R\{a \leqslant x \leqslant b,\ c \leqslant y \leqslant d\}$（如图 6.3 所示）域上采用辛卜生公式求取(6.108)的积分近似值时，可令

$$x_0 = a,\ x_1 = a + h,\ x_2 = a + 2h = b$$
$$y_0 = c,\ y_1 = c + k,\ y_2 = c + 2k = d$$

其中

$$h = \frac{b-a}{2},\ k = \frac{d-c}{2}$$

这时

$$\begin{aligned} \int_a^b \int_c^d f(x, y)\mathrm{d}x\mathrm{d}y &\approx \int_c^d \frac{h}{3}[f(x_0, y) + 4f(x_1, y) + f(x_2, y)]\mathrm{d}y \\ &= \frac{h}{3}\Big\{\frac{k}{3}[f(x_0, y_0) + 4f(x_0, y_1) + f(x_0, y_2)] + \\ &\quad \frac{4k}{3}[f(x_1, y_0) + 4f(x_1, y_1) + f(x_1, y_2)] + \\ &\quad \frac{k}{3}[f(x_2, y_0) + 4f(x_2, y_1) + f(x_2, y_2)]\Big\} \\ &= \frac{hk}{9}[(f_{00} + f_{20} + f_{02} + f_{22}) + 4(f_{10} + f_{01} + f_{21} + f_{12}) + 16f_{11}] \end{aligned}$$

$$=\frac{hk}{9}[\sigma_0+4\sigma_1+16\sigma_2] \tag{6.110}$$

式中，$\sigma_0$ 为矩形顶点上函数值之和；$\sigma_1$ 为矩形各边中点上的函数值之和；$\sigma_2$ 为矩形中心点上的函数值。

各节点相应的系数值 $\lambda_{ij}$ 如图 6.3 所示。

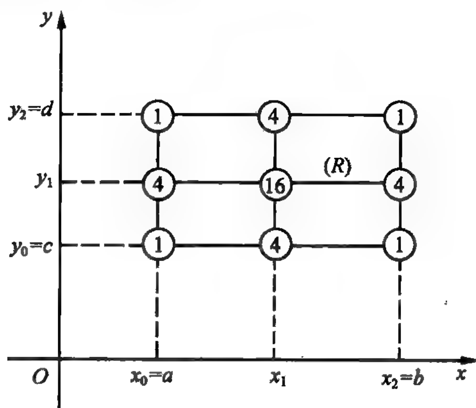



图 6.3

如将矩形域 $R$ 作进一步细分，取

$$h=\frac{b-a}{2n},\ k=\frac{d-c}{2m}$$

相应的节点有

$$\begin{cases} x_i=x_0+ih & (i=0,1,2,\cdots,2n) \\ y_j=y_0+jk & (j=0,1,2,\cdots,2m) \end{cases}$$

则 $R$ 由矩形块 $R_{ij}\,(i=0,1,2,\cdots,n-1;\ j=0,1,2,\cdots,m-1)$ 所组成（图 6.4），对于每个矩形块使用公式(6.110)计算并相加得

$$\int_a^b\int_c^d f(x,\ y)\mathrm{d}x\mathrm{d}y \approx \frac{hk}{9}\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}[(f_{2i,2i}+f_{2i+2,2j}+f_{2i,2j+2}+f_{2i+2,2j+2})+ \\ 4(f_{2i+1,2j}+f_{2i,2j+1}+f_{2i+2,2j+1}+f_{2i+1,2j+2})+16f_{2i+1,2j+1} \tag{6.111}$$

各节点相应的系数值 $\lambda_{ij}$ 可用下面矩阵中的对应元素表示

$$\Lambda=\begin{bmatrix} 1 & 4 & 2 & 4 & 2 & \cdots & 4 & 2 & 4 & 1 \\ 4 & 16 & 8 & 16 & 8 & \cdots & 16 & 8 & 16 & 4 \\ 2 & 8 & 4 & 8 & 4 & \cdots & 8 & 4 & 8 & 2 \\ 4 & 16 & 8 & 16 & 8 & \cdots & 16 & 8 & 16 & 4 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 2 & 8 & 4 & 8 & 4 & \cdots & 8 & 4 & 8 & 2 \\ 4 & 16 & 8 & 16 & 8 & \cdots & 16 & 8 & 16 & 4 \\ 1 & 4 & 2 & 4 & 2 & \cdots & 4 & 2 & 4 & 1 \end{bmatrix}$$

如对 $x$ 和 $y$ 分别用 $n$ 个节点和 $m$ 个节点的高斯求积公式时则可得

$$\int_{-1}^{1}\int_{-1}^{1} f(x, y)\mathrm{d}x\mathrm{d}y \approx \sum_{i=1}^{n}\sum_{j=1}^{m} w_i w_j f(x_i, y_j) \tag{6.112}$$

式中，$x_i(i=1,2,\cdots,n)$和 $y_j(j=1,2,\cdots,m)$为高斯求积公式的节点值，$w_i$、$w_j$ 为其相应的系数。

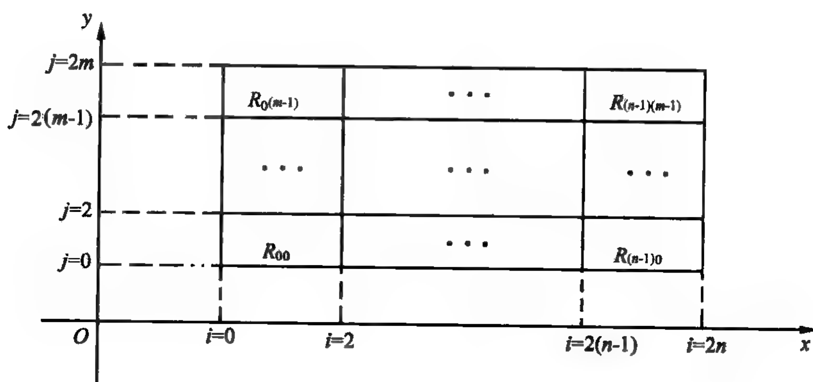



图 6.4

# §2 数值微分

当函数 $f(x)$由表格形式给出时，要寻找 $f(x)$的导数通常称为数值微分。数值微分在目标运动状态的预测分析以及在微分方程的数值方法中具有重要的应用。

## 2.1 差商型数值微分

在微积分中，函数的导数是通过微商的极限定义的

$$f'(x)=\begin{cases}\lim\limits_{h\to 0}\dfrac{f(x+h)-f(x)}{h}\\[2ex] \lim\limits_{h\to 0}\dfrac{f(x)-f(x-h)}{h}\\[2ex] \lim\limits_{h\to 0}\dfrac{f\left(x+\frac{h}{2}\right)-f\left(x-\frac{h}{2}\right)}{h}\end{cases} \tag{6.113}$$

显然，取其达到极限以前的形式就得到了导数的差商近似式

$$f'(x)\approx\begin{cases}\dfrac{f(x+h)-f(x)}{h}\text{（向前差商数值微分公式）}\\[2ex] \dfrac{f(x)-f(x-h)}{h}\text{（向后差商数值微分公式）}\\[2ex] \dfrac{f\left(x+\frac{h}{2}\right)-f\left(x-\frac{h}{2}\right)}{h}\text{（中心差商数值微分公式）}\end{cases} \tag{6.114}$$

从几何上讲，就相当于用弧段的内接弦的斜率代替切线的斜率。由台劳展开式

$$f(x_0+h)=f(x_0)+hf'(x_0)+\frac{h^2}{2}f''(x_0+\theta h),(0<\theta<1)$$

得到式(6.114)$_1$ 的误差为

$$f'(x_0)-\frac{f(x_0+h)-f(x_0)}{h}=-\frac{h}{2}f''(x_0+\theta h)=O(h)$$

同法可得式(6.114)$_2$ 的误差为

$$f'(x_0)-\frac{f(x_0)-f(x_0-h)}{h}=\frac{h}{2}f''(x_0-\theta h)=O(h)$$

至于式(6.114)$_3$ 的误差可由以下两式

$$f\left(x_0+\frac{h}{2}\right)=f(x_0)+\frac{h}{2}f'(x_0)+\frac{1}{2!}\left(\frac{h}{2}\right)^2 f''(x_0)+\frac{1}{3!}\left(\frac{h}{2}\right)^3 f'''\left(x_0+\theta_1\frac{h}{2}\right),(0<\theta_1<1)$$

$$f\left(x_0-\frac{h}{2}\right)=f(x_0)-\frac{h}{2}f'(x_0)+\frac{1}{2!}\left(\frac{h}{2}\right)^2 f''(x_0)-\frac{1}{3!}\left(\frac{h}{2}\right)^3 f'''\left(x_0-\theta_2\frac{h}{2}\right),(0<\theta_2<1)$$

相减得

$$\begin{aligned}f'(x_0)-\frac{f\left(x_0+\frac{h}{2}\right)-f\left(x_0-\frac{h}{2}\right)}{h}&=\frac{1}{48}h^2\left[f'''\left(x_0+\theta_1\frac{h}{2}\right)+f'''\left(x_0-\theta_2\frac{h}{2}\right)\right]\\&=\frac{1}{48}h^2\cdot 2f'''(\xi),\left(x_0-\frac{h}{2}<\xi<x_0+\frac{h}{2}\right)\\&=\frac{1}{24}h^2 f'''(\xi)=O(h^2)\end{aligned}$$

可见中心差商数值微分公式的精度较前面两个数值微分公式为高。从几何上看，弧段内接弦的斜率与切线斜率的平行程度在中点优于两端点，因此用两点的差商值作为其中点处的导数值将会有较高的逼近阶。综上可得以下带有余式的差商型数值微分公式

$$f'(x)\approx\begin{cases}\dfrac{f(x_0+h)-f(x_0)}{h}-\dfrac{h}{2}f''(x_0+\theta h),(0<\theta<1)\\[2ex]\dfrac{f(x_0)-f(x_0-h)}{h}+\dfrac{h}{2}f''(x_0-\theta h),(0<\theta<1)\\[2ex]\dfrac{f\left(x_0+\frac{h}{2}\right)-f\left(x_0-\frac{h}{2}\right)}{h}+\dfrac{1}{24}h^2 f'''(\xi),\left(x_0-\dfrac{h}{2}<\xi<x_0+\dfrac{h}{2}\right)\end{cases}\tag{6.115}$$

由上可见，差商近似微商的误差除取决于函数本身的解析性质外，还取决于所取 $h$ 的大小。$h$ 越小，则误差也越小。但是，太小的 $h$ 将会在计算差商时带来较大的舍入误差，这是由于差商的分子部分相减，二值很接近时有效数字严重损失；又当 $h$ 很小，用 $h$ 作分母时，其除法将会把上述的舍入误差放大，所以必须适当选取 $h$。通常，我们可用事后估计误差法选取步长，记 $R(h)$、$R\left(\frac{h}{2}\right)$为步长取 $h$、$\frac{h}{2}$时的差商型数值微分公式的余式，对于给定的精度要求 $\varepsilon$，当 $\left|R(h)-R\left(\frac{h}{2}\right)\right|<\varepsilon$ 时，步长$\frac{h}{2}$就是合适的步长。

**例 6.11** 用中心差商数值微分公式计算 $f(x)=\sqrt{x}$在 $x=2$ 处的一阶导数值。

**解** 采用计算公式

$$f'(2)\approx\frac{\sqrt{2+h/2}-\sqrt{2-h/2}}{h}=F(2)\tag{6.116}$$

取不同的 $h$ 值按四位小数计算得表 6.7。[$f'(2)$的准确值为 0.353 553]

表 6.7

| $h$ | 2 | 1 | 0.2 | 0.1 | 0.02 | 0.01 | 0.001 |
|---|---|---|---|---|---|---|---|
| $F(2)$ | 0.366 0 | 0.356 4 | 0.353 5 | 0.353 0 | 0.355 0 | 0.350 0 | 0.300 0 |
| 逼近误差 | −0.012 4 | −0.002 8 | −0.000 1 | −0.000 5 | −0.001 4 | 0.003 6 | 0.053 6 |

由表 6.7 可见，$h=0.2$ 时的逼近效果最佳，当 $h$ 再缩小，其逼近效果越来越差。当然对式(6.116)作以下变换

$$f'(2)\approx\frac{\sqrt{2+h/2}-\sqrt{2-h/2}}{h}=\frac{1}{\sqrt{2+h/2}+\sqrt{2-h/2}} \tag{6.117}$$

后再进行计算就可减少有效数字的损失。例如取 $h=0.1$ 时得

$$f'(2)\approx\frac{1}{\sqrt{2+0.05}+\sqrt{2-0.05}}\approx\frac{1}{1.431\,7+1.396\,4}=0.353\,6$$

与准确值比较，它具有四位有效数字，而 $f'(2)\approx0.366\,0$ 只具有两位有效数字。

因函数 $f(x)$ 的导数与其等变元差商间存在以下关系

$$\begin{aligned}
&f'(x)=\frac{\mathrm{d}}{\mathrm{d}x}f(x)=f[x,x]\\
&f''(x)=\frac{\mathrm{d}}{\mathrm{d}x}f(x,x)=2!\ f[x,x,x]\\
&f'''(x)=2!\ \frac{\mathrm{d}}{\mathrm{d}x}f(x,x,x)=3!\ f[x,x,x,x]\\
&\cdots\\
&f^{(k)}(x)=(k-1)!\ \frac{\mathrm{d}}{\mathrm{d}x}f\underbrace{[x,x,\cdots,x]}_{k个}=k!\ f\underbrace{[x,x,\cdots,x]}_{k+1个}
\end{aligned} \tag{6.118}$$

若采用 $n$ 次插值多项式 $P_n(x)\approx f(x)$，则 $f(x)$ 的各阶导数可用 $P_n(x)$ 的各阶等变元差商来近似

$$\begin{aligned}
&f'(x)\approx P_n[x,x]\\
&f''(x)\approx2!\ P_n[x,x,x]\\
&f'''(x)\approx3!\ P_n[x,x,x,x]\\
&\cdots\\
&f^{(k)}(x)\approx k!\ P_n\underbrace{[x,x,\cdots,x]}_{k+1个}
\end{aligned} \tag{6.119}$$

由式(6.118)还可以获得等变元差商的导数与等变元差商间的关系式

$$\begin{aligned}
&f'(x)=f[x,x]\\
&f'(x,x)=2f[x,x,x]\\
&f'(x,x,x)=3f[x,x,x,x]\\
&\cdots\\
&f'\underbrace{[x,x,\cdots,x]}_{k个}=kf\underbrace{[x,x,\cdots,x]}_{k+1个}
\end{aligned} \tag{6.120}$$

**例 6.12**　设有下面的表格函数

| $x$ | 1 | 2 | 4 | 8 | 10 |
|---|---|---|---|---|---|
| $f(x)$ | 0 | 1 | 5 | 21 | 27 |

要求计算 $f'(4)$和 $f''(4)$的近似值 $P_4[4,4]$和 $2P_4[4,4,4]$。

**解** 由于差商具有对称性，我们可将节点 $x=4$ 置于表末，并额外增添 2 个 $x=4$ 的节点，然后建立差商表 6.8。以 $P_4(x)$为 $f(x)$的 4 次插值多项式，其 4 阶差商为常量，所以应有

$$P_4[2,8,10,4,4]=P_4[8,10,4,4,4]=-\frac{1}{144}$$

表 6.8

| $x$ | $y$ | 一阶差商 | 二阶差商 | 三阶差商 | 四阶差商 |
|---|---|---|---|---|---|
| 1 | 0 | | | | |
| | | 1 | | | |
| 2 | 1 | | $\frac{1}{3}$ | | |
| | | $\frac{10}{3}$ | | $-\frac{1}{24}$ | |
| 8 | 21 | | $-\frac{1}{24}$ | | $-\frac{1}{144}$ |
| | | 3 | | $-\frac{1}{16}$ | |
| 10 | 27 | | $-\frac{1}{6}$ | | $P_4[2,8,10,4,4]$ |
| | | $\frac{11}{3}$ | | $P_4[8,10,4,4]$ | |
| 4 | 5 | | $P_4[10,4,4]$ | | $P_4[8,10,4,4,4]$ |
| | | $P_4[4,4]$ | | $P_4[10,4,4,4]$ | |
| 4 | $P_4(4)$ | | $P_4[4,4,4]$ | | |
| | | $P_4[4,4]$ | | | |
| 4 | $P_4(4)$ | | | | |

按差商定义可逐次计算出下面各量

$$P_4[8,10,4,4]=-\frac{1}{16}+(4-2)P_4[2,8,10,4,4]=-\frac{11}{144}$$

$$P_4[10,4,4]=-\frac{1}{6}+(4-8)P_4[8,10,4,4]=\frac{5}{36}$$

$$P_4[4,4]=\frac{11}{3}+(4-10)P_4[10,4,4]=\frac{17}{6}$$

$$P_4[10,4,4,4]=P_4[8,10,4,4]+(4-8)P_4[8,10,4,4,4]=-\frac{7}{144}$$

$$P_4[4,4,4]=P_4[10,4,4]+(4-10)P_4[10,4,4,4]=\frac{31}{72}$$

所以 $f'(4)\approx P_4[4,4]=\frac{17}{6}=2.833$

$$f''(4)\approx 2P_4[4,4,4]=\frac{31}{36}=0.861$$

## 2.2　外推算法求数值微分

对于已建立的数值微分公式，便可运用缩小步长的外推算法，逐次精确地求取导数值。例如，我们运用中心差商数值微分公式来计算导数值

$$f'(x)\approx F(h)=\frac{f\left(x+\frac{h}{2}\right)-f\left(x-\frac{h}{2}\right)}{h},R(h)=\frac{1}{24}h^2f'''(\xi)$$

它的余式与梯形公式的余式相似，因此可仿龙贝格法，类似地建立以下以中心差商数值微分公式为基础的加速收敛序列

$$\begin{cases}T_0^{(0)}=F(h),T_n^{(0)}=F\left(\dfrac{h}{2^n}\right)\\ T_{n-j}^{(j)}=\dfrac{4^jT_{n-j+1}^{(j-1)}-T_{n-j}^{(j-1)}}{4^j-1}\quad(j=1,2,\cdots,n)\end{cases}\tag{6.121}$$

其计算顺序如表 6.9 所示。

**表 6.9**

| $n$ | $T_n^{(0)}$ | $T_{n-1}^{(1)}$ | $T_{n-2}^{(2)}$ |
|---|---|---|---|
| 0 | $T_0^{(0)}=F(h)$ | | |
| 1 | $T_1^{(0)}=F\left(\frac{h}{2}\right)$ | $T_0^{(1)}=\frac{4T_1^{(0)}-T_0^{(0)}}{4-1}$ | |
| 2 | $T_2^{(0)}=F\left(\frac{h}{2^2}\right)$ | $T_1^{(1)}=\frac{4T_2^{(0)}-T_1^{(0)}}{4-1}$ | $T_0^{(2)}=\frac{4^2T_1^{(1)}-T_0^{(1)}}{4^2-1}$ |
| 3 | $T_3^{(0)}=F\left(\frac{h}{2^3}\right)$ | $T_2^{(1)}=\frac{4T_3^{(0)}-T_2^{(0)}}{4-1}$ | $T_1^{(2)}=\frac{4^2T_2^{(1)}-T_1^{(1)}}{4^2-1}$ |
| 4 | $T_4^{(0)}=F\left(\frac{h}{2^4}\right)$ | $T_3^{(1)}=\frac{4T_4^{(0)}-T_3^{(0)}}{4-1}$ | $T_2^{(2)}=\frac{4^2T_3^{(1)}-T_2^{(1)}}{4^2-1}$ |
| ⋮ | ⋮ | ⋮ | ⋮ |

**例 6.13**　使用公式(6.121)计算 $f(x)=\sqrt{x}$ 在 $x=6$ 处的导数值。

**解**　首先使用以下中心差商数值微分公式

$$F(h)=\frac{\sqrt{6+h/2}-\sqrt{6-h/2}}{h}$$

对 $h=2,1,0.5,0.25$ 计算得

$$F(2)=\frac{\sqrt{6+1}-\sqrt{6-1}}{2}=0.204\ 841\ 666$$

$$F(1)=\frac{\sqrt{6+0.5}-\sqrt{6-0.5}}{2}=0.204\ 301\ 876$$

$$F(0.5)=\frac{\sqrt{6+0.25}-\sqrt{6-0.25}}{0.5}=0.204\ 168\ 476$$

$$F(0.25)=\frac{\sqrt{6+0.125}-\sqrt{6-0.125}}{0.25}=0.204\ 135\ 221$$

在此基础上，按式(6.121)建立表6.10，已知 $f'(6)$ 的精确值为0.204 124 145，由表6.10可见，经过三次加速的结果获得的0.204 124 146已具有8位有效数字，可见加速的效果是很显著的。

**表 6.10**

| $n$ | $T_n^{(0)}$ | $T_{n-1}^{(1)}$ | $T_{n-2}^{(2)}$ | $T_{n-3}^{(3)}$ |
|---|---|---|---|---|
| 0 | 0.204 841 666 | | | |
| 1 | 0.204 301 876 | 0.204 121 946 | | |
| 2 | 0.204 168 476 | 0.204 124 009 | 0.204 124 146 | |
| 3 | 0.204 135 221 | 0.204 124 136 | 0.204 124 144 | 0.204 124 144 |

## 2.3 差分型数值微分

为了建立等距节点下的导数计算公式，可对菱形图5.6中的 $C_{t+j}^{i}$ 换以 $(C_{t+j}^{i})'_{t=k}$，便得到在 $x_k$ 点的一阶导数菱形图。如将 $C_{t+j}^{i}$ 换以 $(C_{t+j}^{i})''_{t=k}$，便得到在 $x_k$ 点的二阶导数菱形图。如此做下去，可得到在 $x_k$ 点的各阶导数菱形图，再套用图5.6的使用规则，就可建立插值多项式在 $t=k$ 处的导数 $P_n^{(l)}(x_0+th)|_{t=k}(l=1,2,\cdots)$，最后得

$$\begin{cases} t=\dfrac{x-x_0}{h} \\ P_n^{(l)}(x_k)=\dfrac{1}{h^l}P_n^{(l)}(x_0+th)\Big|_{t=k} \end{cases} \tag{6.122}$$

为示例起见，我们列出了在 $x_0$ 处的一阶导数菱形图6.5及二阶导数菱形图6.6，根据它们不难建立各种插值多项式在 $x_0$ 点的一阶导数或二阶导数公式，如

$$P_n'(x_0)=\frac{1}{h}\left(\Delta y_0-\frac{1}{2}\Delta^2 y_0+\frac{1}{3}\Delta^3 y_0-\frac{1}{4}\Delta^4 y_0+\frac{1}{5}\Delta^5 y_0+\cdots\right) \tag{6.123}$$

$$P_n''(x_0)=\frac{1}{h^2}\left(\Delta^2 y_0-\Delta^3 y_0+\frac{11}{12}\Delta^4 y_0-\frac{5}{6}\Delta^2 y_0+\cdots\right) \tag{6.124}$$

## 2.4 函数值加权型数值微分

如果给定了函数 $f(x)$ 在 $x_0,x_1,\cdots,x_n$ 上的函数值 $y_0,y_1,\cdots,y_n$，则可采用拉格朗日插值多项式的各阶导数值来近似函数的各阶导数值

$$f^{(k)}(x)\approx L_n^{(k)}(x)=\sum_{i=0}^{n}a_i^{(k)}(x)\cdot f(x_i) \tag{6.125}$$

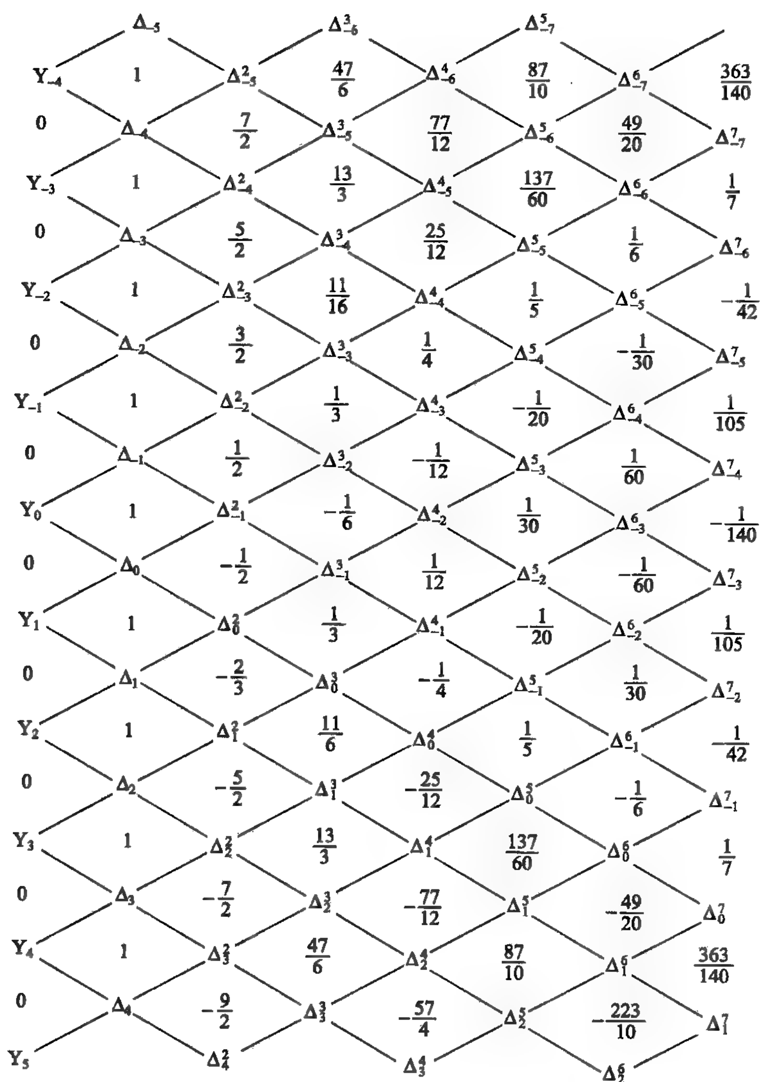



图 6.5

$L_n^{(k)}(x)(k=1,2,\cdots,n)$的余式为

$$
\begin{aligned}
R_n'(x) &= \{\Pi_{n+1}(x)\cdot f[x,x_0,x_1,\cdots,x_n]\}' \\
&= \Pi_{n+1}'(x)\cdot f[x,x_0,x_1,\cdots,x_n]+\Pi_{n+1}(x)\cdot f'[x,x_0,x_1,\cdots,x_n] \\
&= \Pi_{n+1}'(x)\cdot f[x,x_0,x_1,\cdots,x_n]+\Pi_{n+1}(x)\cdot f[x,x,x_0,x_1,\cdots,x_n]
\end{aligned}
$$

$$
\begin{aligned}
R_n''(x) &= \Pi_{n+1}''(x)\cdot f[x,x_0,x_1,\cdots,x_n]+2\Pi_{n+1}'(x)\cdot f[x,x,x_0,x_1,\cdots,x_n]+ \\
&\quad \Pi_{n+1}(x)\cdot f'[x,x,x_0,x_1,\cdots,x_n] \\
&= \Pi_{n+1}''(x)\cdot f[x,x_0,x_1,\cdots,x_n]+2\Pi_{n+1}'(x)\cdot f[x,x,x_0,x_1,\cdots,x_n]+ \\
&\quad \Pi_{n+1}(x)\cdot 2f[x,x,x_1,x_0,x_1,\cdots,x_n]
\end{aligned}
$$

一般有

$$
R_n^{(k)}(x)=\sum_{i=0}^{k}C_k^i f^{(i)}[x,x_0,x_1,\cdots,x_n]\cdot\prod\nolimits_{n+1}^{(k-i)}(x)
$$

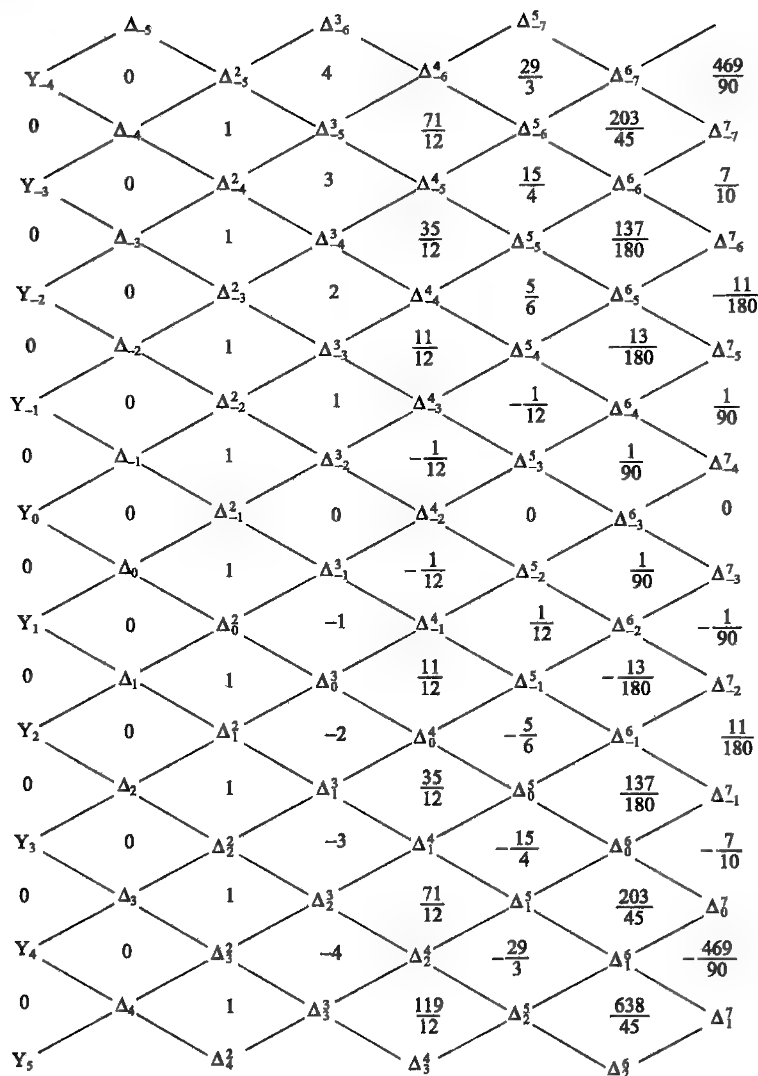



图 6.6

因
$$f^{(i)}[x,x_0,x_1,\cdots,x_n]=\{f'[x,x_0,x_1,\cdots,x_n]\}^{(i-1)}$$
$$=f^{(i-1)}[x,x,x_0,x_1,\cdots,x_n]=\{f'[x,x,x_0,x_1,\cdots,x_n]\}^{(i-2)}$$
$$=2f^{(i-2)}[x,x,x,x_0,x_1,\cdots,x_n]=2\{f'[x,x,x,x_0,x_1,\cdots,x_n]\}^{(i-3)}$$
$$=2\cdot\{3f[x,x,x,x,x_0,x_1,\cdots,x_n]\}^{(i-3)}$$
$$=3!\ f^{(i-3)}[x,x,x,x,x_0,x_1,\cdots,x_n]$$
$$\cdots$$
$$=i!\ f[\underbrace{x,x,\cdots,x}_{i+1个},x_0,x_1,\cdots,x_n]$$

所以
$$R_n^{(k)}(x)=\sum_{i=0}^{k}\frac{k!}{i!(k-i)!}\cdot i!f[\underbrace{x,x,\cdots,x}_{i+1个},x_0,x_1,\cdots x_n]\cdot\prod\nolimits_{n+1}^{(k-i)}(x)$$

$$= \sum_{i=0}^{k} \frac{k!}{(k-i)!} \cdot \frac{f^{(n+i+1)}(\xi_i)}{(n+i+1)!} \cdot \prod_{n+1}^{(k-i)}(x) \tag{6.126}$$

其中　$\min\{x, x_0, x_1, \cdots, x_n\} < \xi_i < \max\{x, x_0, x_1, \cdots, x_n\}$。

由于插值公式的唯一性，式(6.126)不仅适用于 $L_n(x)$ 的情况，它对任何其他 $n$ 次插值多项式均适用。

在等距节点情况下有 $x = x_0 + th, \mathrm{d}x = h\mathrm{d}t$，则

$$f^{(k)}(x) \approx L_n^{(k)}(x_0 + th) = \frac{1}{h^k} \sum_{i=0}^{k} \frac{(-1)^{n-i} y_i}{i!\ (n-i)!} \frac{\mathrm{d}^k}{\mathrm{d}x^k} \frac{t^{[n+1]}}{t-i} \tag{6.127}$$

以下列出等距节点(间距为 $h$)下 $n=1,2,3,4, k=1,2$ 时的导数公式，以备查阅。

一阶数值微分公式：

(1) 两点式($n=1$)

$$\begin{cases} f'(x_0) = \dfrac{1}{h}(y_1 - y_0) - \dfrac{h}{2} f''(\xi) \\ f'(x_1) = \dfrac{1}{h}(y_1 - y_0) + \dfrac{h}{2} f''(\xi) \end{cases} \tag{6.128}$$

(2) 三点式($n=2$)

$$\begin{cases} f'(x_0) = \dfrac{1}{2h}(-3y_0 + 4y_1 - y_2) + \dfrac{h^2}{3} f'''(\xi) \\ f'(x_1) = \dfrac{1}{2h}(y_2 - y_0) - \dfrac{h^2}{6} f'''(\xi) \\ f'(x_2) = \dfrac{1}{2h}(y_0 - 4y_1 + 3y_2) - \dfrac{h^2}{3} f'''(\xi) \end{cases} \tag{6.129}$$

(3) 四点式($n=3$)

$$\begin{cases} f'(x_0) = \dfrac{1}{6h}(-11y_0 + 18y_1 - 9y_2 + 2y_3) - \dfrac{h^3}{4} f^{(4)}(\xi) \\ f'(x_1) = \dfrac{1}{6h}(-2y_0 - 3y_1 + 6y_2 - y_3) + \dfrac{h^3}{12} f^{(4)}(\xi) \\ f'(x_2) = \dfrac{1}{6h}(y_0 - 6y_1 + 3y_2 + 2y_3) - \dfrac{h^3}{12} f^{(4)}(\xi) \\ f'(x_3) = \dfrac{1}{6h}(-2y_0 + 9y_1 - 18y_2 + 11y_3) + \dfrac{h^3}{4} f^{(4)}(\xi) \end{cases} \tag{6.130}$$

(4) 五点式($n=4$)

$$\begin{cases} f'(x_0) = \dfrac{1}{12h}(-25y_0 + 48y_1 - 36y_2 + 16y_3 - 3y_4) + \dfrac{h^4}{5} f^{(5)}(\xi) \\ f'(x_1) = \dfrac{1}{12h}(-3y_0 - 10y_1 + 18y_2 - 6y_3 + y_4) - \dfrac{h^4}{20} f^{(5)}(\xi) \\ f'(x_2) = \dfrac{1}{12h}(y_0 - 8y_1 + 8y_3 - y_4) + \dfrac{h^4}{30} f^{(5)}(\xi) \\ f'(x_3) = \dfrac{1}{12h}(-y_0 + 6y_1 - 18y_2 + 10y_3 + 3y_4) - \dfrac{h^4}{20} f^{(5)}(\xi) \\ f'(x_4) = \dfrac{1}{12h}(3y_0 - 16y_1 + 36y_2 - 48y_3 + 25y_4) + \dfrac{h^4}{5} f^{(5)}(\xi) \end{cases} \tag{6.131}$$

二阶数值微分公式：

(1) 三点式($n=2$)

$$\begin{cases} f''(x_0)=\frac{1}{h^2}(y_0-2y_1+y_2)-hf'''(\xi_1)+\frac{h^2}{6}f^{(4)}(\xi_2) \\ f''(x_1)=\frac{1}{h^2}(y_0-2y_1+y_2)-\frac{h^2}{12}f^{(4)}(\xi_2) \\ f''(x_2)=\frac{1}{h^2}(y_0-2y_1+y_2)-hf'''(\xi_1)-\frac{h^2}{6}f^{(4)}(\xi_2) \end{cases} \tag{6.132}$$

(2) 四点式($n=3$)

$$\begin{cases} f''(x_0)=\frac{1}{6h^2}(12y_0-30y_1+24y_2-6y_3)+\frac{11}{12}h^2f^{(4)}(\xi_1)-\frac{h^3}{10}f^{(5)}(\xi_2) \\ f''(x_1)=\frac{1}{6h^2}(6y_0-12y_1+6y_2)-\frac{1}{12}h^2f^{(4)}(\xi_1)-\frac{h^3}{30}f^{(5)}(\xi_2) \\ f''(x_2)=\frac{1}{6h^2}(6y_1-12y_2+6y_3)-\frac{1}{12}h^2f^{(4)}(\xi_1)-\frac{h^3}{30}f^{(5)}(\xi_2) \\ f''(x_3)=\frac{1}{6h^2}(-6y_0+24y_1-30y_2+12y_3)+\frac{11}{12}h^2f^{(4)}(\xi_1)-\frac{h^3}{10}f^{(5)}(\xi_2) \end{cases} \tag{6.133}$$

(3) 五点式($n=4$)

$$\begin{cases} f''(x_0)=\frac{1}{24h^2}(70y_0-208y_1+228y_2-112y_3+22y_4)-\frac{5}{6}h^3f^{(5)}(\xi_1)+\frac{h^4}{15}f^{(6)}(\xi_2) \\ f''(x_1)=\frac{1}{24h^2}(22y_0-40y_1+12y_2+8y_3-2y_4)+\frac{1}{12}h^3f^{(5)}(\xi_1)-\frac{h^4}{60}f^{(6)}(\xi_2) \\ f''(x_2)=\frac{1}{24h^2}(-2y_0+32y_1-60y_2+32y_3-2y_4)+\frac{h^4}{90}f^{(6)}(\xi) \\ f''(x_3)=\frac{1}{24h^2}(-2y_0+8y_1+12y_2-40y_3+22y_4)-\frac{1}{12}h^3f^{(5)}(\xi_1)+\frac{1}{60}h^4f^{(6)}(\xi_2) \\ f''(x_4)=\frac{1}{24h^2}(22y_0-112y_1+228y_2-208y_3+70y_4)+\frac{5}{6}h^3f^{(5)}(\xi_1)-\frac{1}{15}h^4f^{(6)}(\xi_2) \end{cases} \tag{6.134}$$

**例 6.14** 已知函数 $y=e^x$ 的下列数据

| $x$ | 2.5 | 2.6 | 2.7 | 2.8 | 2.9 |
|---|---|---|---|---|---|
| $y$ | 12.182 5 | 13.463 7 | 14.879 7 | 16.444 6 | 18.174 1 |

试用两点、三点数值微分公式计算 $x=2.7$ 处的函数的一、二阶导数值。

**解** 取 $h=0.2$ 时

$$f'(2.7)\approx\frac{1}{0.2}(14.8797-12.1825)=13.486 \quad (\text{使用式}(6.128))$$

$$f'(2.7)\approx\frac{1}{2\times0.2}(18.1741-12.1825)=14.979 \quad (\text{使用式}(6.129))$$

$$f''(2.7)\approx\frac{1}{0.2^2}(12.1825-2\times14.8797+18.1741)=14.930 \quad (\text{使用式}(6.132))$$

取 $h=0.1$ 时，相应地有

$$f'(2.7)\approx\frac{1}{0.1}(14.8797-13.4637)=14.160$$

$$f'(2.7)\approx\frac{1}{2\times0.1}(16.4446-13.4637)=14.905$$

$$f''(2.7)\approx\frac{1}{0.1^2}(13.4637-2\times14.8797+16.4446)=14.890$$

这里 $f'(2.7)$ 和 $f''(2.7)$ 的真值都是 14.879 73…。上面计算表明，三点公式比两点公式准确，步长越小越准确。一般情况下，这个结论是对的。但由余式可见，如果高阶导数无界，或舍入误差超过截断误差，这个结论就不对了。

### 2.5　用三次样条函数求数值微分

用三次样条函数的导数近似代替函数的导数，当被插值函数 $f(x)$ 有较好的光滑性时，随着 $n$ 的增加，不仅样条函数无限趋近于函数 $f(x)$，而且样条函数的导数也能无限地趋近 $f'(x)$，这种性质要比使用插值多项式优越得多。因此，利用样条函数求导数的近似值比较可靠些。当然用样条函数求数值微分也有其缺点，即建立样条函数时需要解线性代数方程组，相应的计算量要大一些。

### 2.6　使用数值积分公式求微分

在牛顿—莱伯尼兹公式

$$\int_{x_{k-1}}^{x_{k+1}} f'(x)\mathrm{d}x=f(x)\Big|_{x_{k-1}}^{x_{k+1}}=f(x_{k+1})-f(x_{k-1}) \tag{6.135}$$

中，对于左式应用不同的求积公式就可以获得不同的数值微分公式。例如，采用 $Q_{02}$ 求积公式便得

$$2hf'(x_k)+\frac{1}{3}h^3f''(\xi)=f(x_{k+1})-f(x_{k-1})$$

$$f'(x_k)=\frac{f(x_{k+1})-f(x_{k-1})}{2h}-\frac{1}{6}h^2f''(\xi)\quad(x_{k-1}\leqslant\xi\leqslant x_{k+1}) \tag{6.136}$$

为提高精度，今采用辛卜生公式，就有

$$\frac{h}{3}(f'(x_{k-1})+4f'(x_k)+f'(x_{k+1})-\frac{h^5}{90}f^{(4)}(\xi)=f(x_{k+1})-f(x_{k-1})\quad(x_{k-1}\leqslant\xi\leqslant x_{k+1})$$

略去余式后，上式化为

$$f'(x_{k-1})+4f'(x_k)+f'(x_{k+1})=\frac{3[f(x_{k+1})-f(x_{k-1})]}{h}\ (k=1,2,\cdots,n-1) \tag{6.137}$$

上式中，未知数为 $f'(x_0),f'(x_1),\cdots,f'(x_n)$，方程数为 $n-1$ 个，需附加二个边界条件 $f'(x_0)$ 和 $f'(x_n)$，这样就变成 $n-1$ 个未知数的 $n-1$ 个方程组了，可用矩阵形式将它表为

$$\begin{bmatrix}4&1&&&0\\1&4&1&&\\&\ddots&\ddots&\ddots&\\&&1&4&1\\0&&&1&4\end{bmatrix}\begin{bmatrix}f'(x_1)\\f'(x_2)\\\vdots\\f'(x_{n-2})\\f'(x_{n-1})\end{bmatrix}=\begin{bmatrix}3[f(x_2)-f(x_0)]/h-f'(x_0)\\3[f(x_3)-f(x_1)]/h\\\vdots\\3[f(x_{n-1})-f(x_{n-3})]/h\\3[f(x_n)-f(x_{n-2})]/h-f'(x_n)\end{bmatrix} \tag{6.138}$$

由它可以解出 $f'(x_1),f'(x_2),\cdots,f'(x_{n-1})$ 的数值。

如果 $f''(x_0)$，$f''(x_n)$ 的值已知，可将式(6.138)中的导数阶数均递增一次，便得以下方程组

$$\begin{bmatrix} 4 & 1 & & & 0 \\ 1 & 4 & 1 & & \\ & \ddots & \ddots & \ddots & \\ 0 & & & 1 & 4 \end{bmatrix} \begin{bmatrix} f''(x_1) \\ f''(x_2) \\ \vdots \\ f''(x_{n-1}) \end{bmatrix} = \begin{bmatrix} 3[f'(x_2)-f'(x_0)]/h - f''(x_0) \\ 3[f'(x_3)-f'(x_1)]/h \\ \vdots \\ 3[f'(x_n)-f'(x_{n-2})]/h - f''(x_n) \end{bmatrix} \tag{6.139}$$

由此可以解得二阶导数值 $f''(x_1)$，$f''(x_2)$，$\cdots$，$f''(x_{n-1})$。

在方程组(6.138)求解前，必须增加 2 个条件，可将它设置为边界条件。在边界条件未知的情况下，可以使用差商来近似。例如，若对 $f'(x_1)$、$f'(x_{n-1})$ 使用中心差商数值微分公式的计算值作为 2 个补充条件，则式(6.137)成为

$$\begin{cases} f'(x_1) = \dfrac{f(x_2)-f(x_0)}{2h} \\ f'(x_{k-1}) + 4f'(x_k) + f'(x_{k+1}) = \dfrac{3[f(x_{k+1})-f(x_{k-1})]}{h} \quad (k=1,2,\cdots,n-1) \\ f'(x_{n-1}) = \dfrac{f(x_n)-f(x_{n-2})}{2h} \end{cases} \tag{6.140}$$

由此便可解得 $f'(x_0)$，$f'(x_2)$，$\cdots$，$f'(x_{n-2})$，$f'(x_n)$。

# 习 题 六

6.1　分别用梯形公式和辛卜生公式计算下列积分。

(1) $\int_0^1 \frac{x}{4+x^2}dx$，$M=8$

(2) $\int_0^1 \frac{(1-e^{-x})^{\frac{1}{2}}}{x}dx$，$M=10$

(3) $\int_0^9 \sqrt{x}dx$，$M=4$

6.2　用高斯求积公式计算下列积分的值。

$$\int_0^{\frac{\pi}{2}} \sqrt{1-\frac{1}{2}\sin^2 t}dt$$

6.3　用下列方法计算积分$\int_1^3 \frac{dy}{y}$。

(1) 龙贝格方法；

(2) 用 $n=3$ 及 $n=5$ 的高斯求积公式。

6.4　计算下列积分。

(1) $\int_0^1 \frac{dx}{1+x}$；　(2) $\int_0^2 \frac{dx}{1+x}$；　(3) $\int_0^1 \frac{dx}{1+x^2}$；

(4) $\int_0^{\frac{\pi}{2}} \frac{\sin x}{x}dx$；　(5) $\int_0^1 \frac{\ln(1+x)}{1+x^2}dx$；　(6) $\int_0^1 \frac{\ln(1+x)}{x}dx$

用复化梯形公式、复化辛卜生公式、龙贝格法、高斯求积公式，结果精确到 $10^{-4}$。

6.5　对定积分$\int_{0.5}^1 \sqrt{x}dx$，试用梯形公式、辛卜生公式和龙贝格法进行计算，并研究用辛卜生公式计算所得误差。

6.6　求系数 $A_1, A_2, A_3$，使求积公式

$$\int_{-1}^1 f(x)dx \approx A_1 f(-1) + A_2 f(-\frac{1}{3}) + A_3 f(\frac{1}{3})$$

对于次数≤2 的一切多项式都精确成立。

6.7　用复化辛卜生公式计算积分

$$\iint_R \ln(x+2y)dxdy$$

式中，$R=\{(x,y)|_{1.4\leqslant x\leqslant 2.0,\ 1.0\leqslant y\leqslant 1.5}\}$，取 5 位小数进行计算之。

6.8　给定下列表格值

| $x$ | 50 | 55 | 60 | 65 |
|---|---|---|---|---|
| $y$ | 1.699 0 | 1.740 4 | 1.778 2 | 1.812 9 |

利用四点式($n=3$)求 $f'(50)$、$f''(50)$的近似值。

6.9　计算积分$\int_0^1 e^x dx$，若用复化梯形公式，问区间应分多少等份，才能保证计算结果有 5

位有效数字?

6.10 用复化辛卜生公式求$\int_1^3 e^x \sin x\mathrm{d}x$的近似值,确定 $m$ 和 $h$ 使公式的截断误差小于$10^{-6}$。

6.11 用切比雪夫求积公式计算

$$I=\int_{-1}^{+1}\sin x\mathrm{d}x$$

要求总误差小于 0.002,问公式中应取几项合适,字长至少应取几位?

6.12 要求使用计算量最少的高斯求积公式计算

$$I=\int_2^4 \frac{1}{x}\mathrm{d}x$$

的数值,使其总误差不超过 0.005。

# 第七章　常微分方程数值解法

## §1 引　　言

在科学技术中遇到的常微分方程，只有少数是较简单和典型的常微分方程，比如线性常系数微分方程能够用初等方法求得它们的解析解。对于变系数微分方程的求解就有困难，更不用说一般的非线性微分方程了。多数情况下，常微分方程只能采用近似方法求解。近似方法有两类，一类称为近似解析方法，如级数解法、逐次逼近法等；另一类近似方法称为数值解法，它可以给出解在一些离散点上的近似值。利用电子计算机解常微分方程主要用数值解法。

首先考虑初值问题

$$\begin{cases} y' = f(x,y) \\ y(x_0) = y_0 \end{cases} \tag{7.1}$$

的数值解法，因为这一方法可以推广到一阶微分方程组，而高阶方程又可化为一阶微分方程组，所以我们首先研究它的数值解法。在叙述数值解法之前，我们需要考虑解的存在性。因为如果问题没有解，即使利用数值解法可以求得一些数据也是毫无意义的。另外，在解存在的情况下，必须保证初值问题具有唯一的解。关于解的存在唯一性有如下的定理描述。

**定理 7.1**　设 $f(x,y)$ 在域 $D=\{a\leqslant x\leqslant b, -\infty<y<\infty\}$ 上有定义且连续，同时满足如下的李普希兹条件

$$|f(x,y)-f(x,y^*)|\leqslant L|y-y^*|, \qquad (x,y)\in D,(x,y^*)\in D \tag{7.2}$$

式中，$L$ 为李普希兹常数，则初值问题(7.1)的解存在且唯一，并且解 $y(x)$ 连续可微。

在 $f(x,y)$ 对 $y$ 可微的情况下，若偏导数有界，则可取

$$L = \max_{(x,y)\in D}\left|\frac{\partial f(x,y)}{\partial y}\right| \tag{7.3}$$

这时李普希兹条件显然成立

$$|f(x,y)-f(x,y^*)| = \left|\frac{\partial f(x,y)}{\partial y}(y-y^*)\right| \leqslant L|y-y^*| \tag{7.4}$$

这通常是验证式(7.2)是否满足的最简便的方法。

除了需要保证初值问题有解外，还必须保证微分方程本身是适定的，在微分方程教材中，关于适定的问题有下面的定理描述。

**定理 7.2**　如果 $f(x,y)$ 满足李普希兹条件，则初值问题(7.1)是适定的。

本章主要讨论初值问题(7.1)的常用数值解法，并假定 $f(x,y)$ 满足存在唯一性定理及适当光滑等条件。初值问题(7.1)的解 $y(x)$ 是$[a,b]$上变量 $x$ 的连续函数。求这个初值问题的数值解，就是在区间$[a,b]$上的若干离散点，如在

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

上，用离散化方法将初值问题(7.1)化成离散变量的相应问题，把相应问题的解 $y_k(k=1,2,\cdots,n)$ 作为初值问题(7.1)理论解 $y(x_k)$ 的近似值。

称 $y_k(k=1,2,\cdots,n)$ 为初值问题(7.1)的数值解。记

$$x_{k+1} = x_k + h_k \quad (k = 0,1,2,\cdots,n-1)$$

$h_k$ 称为步长。在等步长 $h_k=h$ 时,得

$$x_k = a + kh \quad (k = 0,1,2,\cdots,n)$$

数值解法的关键在于设法消去初值问题(7.1)中的导数项,这一过程称为"离散化"。通过离散化过程,就可将微分方程转化为差分方程来求解。离散化方法有多种,可基于数值微分近似函数的导数;也可基于台劳级数法及基于台劳展开式的待定系数法以及基于数值积分公式等途径建立差分方程。

# §2 台劳级数法

初值问题(7.1)由初值点$(x_0,y_0)$出发,取步长为 $h$,建立 $x_0$ 点处的 $p$ 阶台劳多项式计算 $x_1=x_0+h$ 点的 $y_1$ 值

$$y_1 = y(x_0) + hy^{(1)}(x_0) + \frac{h^2}{2!}y^{(2)}(x_0) + \cdots + \frac{h^p}{p!}y^{(p)}(x_0) \tag{7.5}$$

其中的各阶导数值计算如下:

$$\begin{cases} y(x_0) = y_0 \\ y'(x_0) = f(x_0,y_0) = f_0 \\ y''(x_0) = \left(\dfrac{\partial f}{\partial x} + \dfrac{\partial f}{\partial y}y'\right)\Big|_{x=x_0} = (f_x' + f_y' f)_0 \\ y'''(x_0) = (f_{xx}'' + 2f f_{xy}'' + f^2 f_{yy}'' + f_x' f_y' + f(f_y')^2)_0 \\ \cdots \end{cases} \tag{7.6}$$

然后以$(x_1,y_1)$为起点,建立 $x_1$ 点的 $p$ 阶台劳多项式计算 $x_2$ 点的 $y_2$ 值,如此类推,就能求得数值解 $y_1,y_2,\cdots y_n$。

为了衡量数值方法的误差,在一步计算中,假定 $y_n$ 为准确值,即 $y_n$ 取理论值 $y(x_n)$,在此前提下(称为局部化假定),用某种数值方法精确计算得到 $y_{n+1}$,则称 $R_{n+1}=y(x_{n+1})-y_{n+1}$ 为该数值方法的局部截断误差。并称

$$\varepsilon_{i+1} = y(x_{i+1}) - y_{i+1} \tag{7.7}$$

为该数值方法的整体截断误差。其中 $y(x_{i+1})$为初值问题(7.1)的理论解,$y_{i+1}$为不计舍入误差时用该数值方法从 $x_0$ 开始,逐步得到在 $x_{i+1}$处的数值解。

局部截断误差与整体截断误差是该数值方法不计舍入误差时的理论误差,而后面讲的稳定性概念则是研究舍入误差对计算结果的影响问题。

**定义 7.1** 若数值方法的局部截断误差为

$$R_{n+1} = O(h^{p+1}) \tag{7.8}$$

则称该方法具有 $p$ 阶精度或 $p$ 阶方法,这里 $p$ 为正整数。

局部截断误差仅是由 $n$ 到 $n+1$ 这一步执行数值方法所引起的误差,它的大小可以反映出该方法精度高低。为了提高公式的精度,则要将 $p$ 值取大,但是,当 $p$ 值取大时,公式变得复杂,计算量就会增大。

采用台劳级数法,只要初值问题的理论解充分光滑,就可获得较高准确度的数值解。但是需计算 $y(x)$的各阶导数,而当 $f(x,y)$的表达式复杂时,求取各阶导数是很烦琐的。一般不直

接采用它来求取数值解，但在下面介绍的多步法中可用它来计算比较精确的起始值。

**例 7.1**　取步长 $h=0.1$，用一、二、四阶台劳多项式求解初值问题

$$y'=y^2,\qquad 0\leqslant x\leqslant \frac{1}{2},\qquad y(0)=1$$

**解**　由于

$$\begin{aligned} y' &= y^2 \\ y'' &= 2yy' = 2y^3 \\ y''' &= 6y^2y' = 6y^4 \\ y^{(4)} &= 24y^3y' = 24y^5 \end{aligned}$$

相应的一、二、四阶台劳多项式为

$$p=1,\qquad y_{n+1}=y_n+hy_n^2=y_n(hy_n+1)$$

$$p=2,\qquad y_{n+1}=y_n+hy_n^2+\frac{h^2}{2!}\cdot 2y_n^3=y_n[(hy_n+1)hy_n+1]$$

$$\begin{aligned} p=4,\qquad y_{n+1} &= y_n+hy_n^2+\frac{h^2}{2!}\cdot 2y_n^3+\frac{h^3}{3!}\cdot 6y_n^4+\frac{h^4}{4!}\cdot 24y_n^5 \\ &= y_n((((hy_n+1)hy_n+1)hy_n+1)hy_n+1) \end{aligned}$$

从 $y_0=1$ 开始，按上列公式计算，得结果如表 7.1 所示，表中最后一行是真解 $y(x_n)=\dfrac{1}{1-x_n}$ 的值。

**表 7.1**

| $p$ \ $y_n$ \ $x_n$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
|---|---|---|---|---|---|
| 1 | 1.100 00 | 1.221 00 | 1.370 08 | 1.557 79 | 1.800 46 |
| 2 | 1.110 00 | 1.246 89 | 1.421 74 | 1.662 62 | 1.970 87 |
| 4 | 1.111 10 | 1.249 66 | 1.428 48 | 1.666 45 | 1.999 42 |
| $y(x_n)$ | 1.111 11 | 1.250 00 | 1.428 57 | 1.666 67 | 2.000 00 |

## §3　基于数值微分公式的方法

对于式(7.1)中的 $y'=f(x,y)$ 使用向前差商数值微分公式(6.1)近似有

$$y'(x_n)\approx\frac{y(x_{n+1})-y(x_n)}{h}$$

代入 $y'(x_n)=f(x_n,y(x_n))$ 左端，得

$$\frac{y(x_{n+1})-y(x_n)}{h}\approx f(x_n,y(x_n))$$

再用 $y(x_n)$ 的近似值 $y_n$ 代入上式得

$$y(x_{n+1})\approx y_n+hf(x_n,y_n)$$

今把上式右端计算出来的值记为 $y_{n+1}$，把它作为 $y(x_{n+1})$ 的近似值，有

$$y_{n+1} = y_n + hf(x_n, y_n) \tag{7.9}$$

称为欧拉(Euler)公式，其局部截断误差为

$$R_{n+1} = \frac{h^2}{2} y''(\xi_n) = O(h^2), \quad x_n < \xi_n < x_{n+1} \tag{7.10}$$

若采用数值微分公式(6.2)的向后差商近似导数时，有

$$y'(x_{n+1}) \approx \frac{y(x_{n+1}) - y(x_n)}{h}$$

就可得以下一个求解公式

$$y_{n+1} = y_n + hf(x_{n+1}, y_{n+1}) \tag{7.11}$$

称为后退欧拉公式，其局部截断误差为

$$R_{n+1} = -\frac{h^2}{2} y''(\tilde{\xi}_n) = O(h^2), \quad x_n < \tilde{\xi} < x_{n+1} \tag{7.12}$$

我们从欧拉公式和后退欧拉公式的局部截断误差式(7.10)和式(7.12)可见，若取式(7.9)与式(7.11)的平均值作为一个求解公式，则可能抵消原公式的误差主部而获得精度更好的公式，这个公式称为梯形公式或改进的欧拉公式

$$y_{n+1} = y_n + \frac{h}{2}[f(x_n, y_n) + f(x_{n+1}, y_{n+1})] \tag{7.13}$$

其局部截断误差(参见表 7.4)为

$$R_{n+1} = -\frac{h^3}{12} y'''(\xi_n) = O(h^3), \quad x_n < \xi_n < x_{n+1} \tag{7.14}$$

**定义 7.2** 在差分方程中，对任何 $f$ 若 $y_{n+1}$ 已被明显地解出来，就称这样的方法为显式法；否则称为隐式法。

欧拉法是显式法，梯形法是隐式法。显式法用递推的数值求解方法。隐式法求解方法用迭代法，每一步先要用一个显式公式为它提供迭代初值，然后用迭代法求解隐式公式。以梯形公式为例，可用欧拉公式计算初值，然后用梯形公式进行迭代计算：

$$\begin{cases} y_{n+1}^{(0)} = y_n + hf(x_n, y_n) \\ y_{n+1}^{(k+1)} = y_n + \dfrac{h}{2}[f(x_n, y_n) + f(x_{n+1}, y_{n+1}^{(k)})] \quad (k = 0, 1, 2, \cdots) \end{cases} \tag{7.15}$$

反复迭代，直到满足精度要求 $|y_{n+1}^{(k+1)} - y_{n+1}^{(k)}| < \varepsilon$ 为止，此时的 $y_{n+1}^{(k+1)}$ 作为 $x_{n+1}$ 点的数值解。

# §4 龙格-库塔法

龙格首先提出了间接使用台劳展开式的方法。具体做法是，用 $f(x,y)$ 在不同点上的函数值的线性组合来代替式(7.5)，即等量置换的方法。下面以二阶龙格-库塔公式为例来说明这一基本思想。

设截止到 $h^2$ 项的台劳展开式为

$$y(x_{n+1}) = y(x_n) + hf(x_n, y(x_n)) + \frac{h^2}{2!}(f_x' + f_y' f) \tag{7.16}$$

假定上式数值等价于

$$y_{n+1} = c_1 k_1 + c_2 k_2 \tag{7.17}$$

其中

$$\begin{cases}k_1=hf(x_n,y_n)\\k_2=hf(x_n+\lambda_1 h,y_n+\mu_1 k_1)\end{cases}$$

式中的参数 $c_1$、$c_2$、$\lambda_1$、$\mu_1$ 待定，要求方法达到二阶精度。为此将 $k_2$ 在 $(x_n,y_n)$ 点展开为

$$k_2=h[f(x_n,y_n)+\lambda_1 hf_x'(x_n,y_n)+\mu_1 hf(x_n,y_n)f_y'(x_n,y_n)]+O(h^3)$$

代入式(7.17)得

$$\begin{aligned}y_{n+1}=&y_n+h(c_1+c_2)f(x_n,y_n)+c_2h^2[\lambda_1 f_x'(x_n,y_n)+\\&\mu_1 f(x_n,y_n)\cdot f_y'(x_n,y_n)]+O(h^3)\end{aligned}\tag{7.18}$$

由于要求方法是二阶的，所以(7.18)与(7.16)关于 $h$ 同次幂的系数必须相等，由此得到

$$\begin{cases}c_1+c_2=1\\c_2\lambda_1=\dfrac{1}{2}\\c_2\mu_1=\dfrac{1}{2}\end{cases}\tag{7.19}$$

四个未知量 $c_1$、$c_2$、$\lambda_1$、$\mu_1$ 只有三个方程，有一个参量可任选，如取 $c_2$ 为自由参量(非零)，于是

$$\begin{cases}c_1=1-c_2\\\lambda_1=\mu_1=\dfrac{1}{2c_2}\end{cases}\tag{7.20}$$

常见的取法为 $c_2=\frac{1}{2}$，则 $c_1=\frac{1}{2}$，$\lambda_1=\mu_1=1$，于是式(7.17)成为

$$\begin{cases}y_{n+1}=y_n+\dfrac{1}{2}(k_1+k_2)\\k_1=hf(x_n,y_n)\\k_2=hf(x_n+h,\ y_n+k_1)\end{cases}\tag{7.21}$$

如果取 $c_2=1$，则 $c_1=0$，$\lambda_1=\mu_1=\frac{1}{2}$，则(7.17)式成为

$$y_{n+1}=y_n+hf(x_n+\frac{1}{2}h,\ y_n+\frac{1}{2}hf(x_n,y_n))\tag{7.22}$$

由于方程组(7.19)为不定方程组，其解不唯一，还有许多可能的取法，相应公式的局部截断误差都是 $O(h^3)$，统称这些公式为二阶龙格-库塔公式。仿照以上方法可以开发出各阶龙格-库塔公式如下。

$p=1$(一阶)：

$$\begin{cases}y_{n+1}=y_n+k_1\\k_1=hf(x_n,y_n)\end{cases}\tag{7.23}$$

$p=2$(二阶)：

$$\begin{cases}y_{n+1}=y_n+k_2\\k_1=hf(x_n,y_n)\\k_2=hf\left(x_n+\dfrac{h}{2},y_n+\dfrac{k_1}{2}\right)\end{cases}\tag{7.24}$$

——修正的欧拉法或中点公式

$$\begin{cases}y_{n+1}=y_n+\dfrac{1}{4}(k_1+3k_2)\\k_1=hf(x_n,y_n)\\k_2=hf\left(x_n+\dfrac{2}{3}h,y_n+\dfrac{2}{3}k_1\right)\end{cases}\tag{7.25}$$

——Heun 法

$$\begin{cases} y_{n+1}=y_n+\frac{1}{2}(k_1+k_2) \\ k_1=hf(x_n,y_n) \\ k_2=hf(x_n+h,y_n+k_1) \end{cases} \tag{7.26}$$

——改进的欧拉法

$p=3$（三阶）：

$$\begin{cases} y_{n+1}=y_n+\frac{1}{4}(k_1+3k_3) \\ k_1=hf(x_n,y_n) \\ k_2=hf\left(x_n+\frac{h}{3},y_n+\frac{k_1}{3}\right) \\ k_3=hf\left(x_n+\frac{2}{3}h,y_n+\frac{2}{3}k_2\right) \end{cases} \tag{7.27}$$

——Heun 三阶公式

$$\begin{cases} y_{n+1}=y_n+\frac{1}{6}(k_1+4k_2+k_3) \\ k_1=hf(x_n,y_n) \\ k_2=hf\left(x_n+\frac{h}{2},y_n+\frac{k_1}{2}\right) \\ k_3=hf(x_n+h,y_n-k_1+2k_2) \end{cases} \tag{7.28}$$

——库塔三级算法

$$\begin{cases} y_{n+1}=y_n+\frac{1}{9}(2k_1+3k_2+4k_3) \\ k_1=hf(x_n,y_n) \\ k_2=hf\left(x_n+\frac{h}{2},y_n+\frac{k_1}{2}\right) \\ k_3=hf\left(x_n+\frac{3}{4}h,y_n+\frac{3}{4}k_2\right) \end{cases} \tag{7.29}$$

$p=4$（四阶）：

$$\begin{cases} y_{n+1}=y_n+\frac{1}{6}(k_1+2k_2+2k_3+k_4) \\ k_1=hf(x_n,y_n) \\ k_2=hf\left(x_n+\frac{h}{2},y_n+\frac{k_1}{2}\right) \\ k_3=hf\left(x_n+\frac{h}{2},y_n+\frac{k_2}{2}\right) \\ k_4=hf(x_n+h,y_n+k_3) \end{cases} \tag{7.30}$$

——古典公式

$$
\begin{cases}
y_{n+1}=y_n+\dfrac{1}{8}(k_1+3k_2+3k_3+k_4)\\
k_1=hf(x_n,y_n)\\
k_2=hf\left(x_n+\dfrac{h}{3},y_n+\dfrac{k_1}{3}\right)\\
k_3=hf\left(x_n+\dfrac{2}{3}h,y_n+\dfrac{k_1}{3}+k_2\right)\\
k_4=hf(x_n+h,\ y_n+k_1-k_2+k_3)
\end{cases}
\tag{7.31}
$$

——库塔公式

$$
\begin{cases}
y_{n+1}=y_n+\dfrac{1}{6}[k_1+(2-\sqrt{2})k_2+(2+\sqrt{2})k_3+k_4]\\
k_1=hf(x_n,y_n)\\
k_2=hf\left(x_n+\dfrac{h}{2},y_n+\dfrac{k_1}{2}\right)\\
k_3=hf\left(x_n+\dfrac{h}{2},y_n+\dfrac{\sqrt{2}-1}{2}k_1+\dfrac{2-\sqrt{2}}{2}k_2\right)\\
k_4=hf(x_n+h,\ y_n-\dfrac{\sqrt{2}}{2}k_2+\dfrac{2+\sqrt{2}}{2}k_3)
\end{cases}
\tag{7.32}
$$

——Gill 公式

我们还可以造出很多具体的计算公式，在龙格-库塔公式中所需计算 $f(x,y)$ 的次数叫做它的级数，公式中所能达到的 $h$ 的最高方次称为它的阶数，它们之间的关系由 Butcher(1965年)给出，如表 7.2 所示。

表 7.2

| $N$(计算 $f(x,y)$ 的次数) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$(方法的最大阶数) | 1 | 2 | 3 | 4 | 4 | 5 | 6 | 6 | 7 | 8 | 9 |

当 $N\geqslant 10$ 时，$p\leqslant N-2$。由上可见，四级以下的龙格-库塔公式的级数与阶数一致，但是 $R>4$ 级以上的龙格-库塔公式就不一定是 $R$ 阶的了。

龙格-库塔法是常用的有效方法，它的优点是：① 方法简练，易编程序；② 龙格-库塔法只需知道起点值即可逐步以定步长或变步长外推，称这种方法为单步法；③ 具有良好的数值稳定性，即随着计算步数的增大，因初始误差或舍入误差的影响导致数值解的误差不会增长或误差有界。缺点是每一步的计算量较大；公式的局部截断误差难以求得。从计算量着眼，一般低精度问题宜采用低阶龙格-库塔公式，而高精度问题以采用高阶龙格-库塔为宜。至于四阶以上较少被采用。

为了保证龙格-库塔公式的局部截断误差满足精度要求 $\varepsilon$ ，必须选定合适的步长 $h$，我们以四阶龙格-库塔法为例叙述其确定方法。

今从某一点 $x_n$ 出发，初定 $h$ 为步长，经过一步计算得 $y(x_{n+1})$ 的近似值 $y_{n+1}^{(h)}$，其局部截断误差为

$$
y(x_{n+1})-y_{n+1}^{(h)}\approx ch^5 \tag{7.33}
$$

当 $h$ 不大时，$c$ 可近似地看做常数。然后将步长折半，即取 $h/2$ 为步长，从 $x_n$ 出发经两步计算

求得 $y(x_{n+1})$ 的近似值为 $y_{n+1}^{(\frac{h}{2})}$，其每一步的局部截断误差为 $c(\frac{h}{2})^5$，于是有

$$y(x_{n+1})-y_{n+1}^{(\frac{h}{2})}\approx 2c(\frac{h}{2})^5 \tag{7.34}$$

式(7.34)与式(7.33)相比可得

$$\frac{y(x_{n+1})-y_{n+1}^{(\frac{h}{2})}}{y(x_{n+1})-y_{n+1}^{(h)}}\approx\frac{1}{16}$$

$$y(x_{n+1})-y_{n+1}^{(\frac{h}{2})}\approx\frac{1}{15}(y_{n+1}^{(\frac{h}{2})}-y_{n+1}^{(h)})$$

这表明以 $y_{n+1}^{(\frac{h}{2})}$ 作为 $y(x_{n+1})$ 的近似值，其误差可用先后两次计算结果之差来表示。因而只需考察

$$\frac{1}{15}|y_{n+1}^{(\frac{h}{2})}-y_{n+1}^{(h)}|<\varepsilon \tag{7.35}$$

或

$$|y_{n+1}^{(\frac{h}{2})}-y_{n+1}^{(h)}|<\varepsilon \tag{7.36}$$

是否成立。上述两个不等式统一写为

$$\Delta<\varepsilon \tag{7.37}$$

是否成立。若成立，将 $h$ 加倍至 $2h$ 再作一次，若仍然成立，则将 $2h$ 再加倍，直到不再小于 $\varepsilon$ 为止，退回一步即可得到所要求的步长 $h$。若 $\Delta>\varepsilon$，则将 $h$ 折半为 $h/2$ 再作，如此类推直到式(7.37)成立为止，并取最后的 $h/2$ 为所需的步长。

上述确定步长的方法亦可用来设计变步长龙格-库塔法之用。变步长龙格-库塔法进行如下，在计算过程的第一步，比如从 $x_n$ 出发，以步长 $h$ 计算出 $y_{n+1}^{(h)}$，再以 $h/2$ 步长经两步计算出 $y_{n+1}^{(\frac{h}{2})}$，判 $\Delta<\varepsilon$? 以下区分两种情况处理。

① 如果 $\Delta>\varepsilon$，我们反复将步长折半进行计算，直到 $\Delta<\varepsilon$ 为止。这时取最终得到的 $y_{n+1}^{(\frac{h}{2})}$ 作为结果。

② 如果 $\Delta<\varepsilon$，我们反复将步长加倍进行计算，直到 $\Delta>\varepsilon$ 为止，这时再将步长折半一次就得到所要的结果。

这种通过加倍或折半的手续处理步长的方法称作变步长方法。表面上，为了选取步长，每一步中的计算量增加了，但从整体上看，这种变步长的方法是合算的。

**例 7.2** 用四阶古典龙格-库塔法解初值问题

$$\begin{cases} y'=x-y+1, & 0\leqslant x\leqslant 1 \\ y(0)=1 \end{cases}$$

**解** 取步长 $h=0.1$，按式(7.30)计算如下。

$k_1=0.1\times(0-1+1)=0$

$k_2=0.1\times[(0+\frac{0.1}{2})-(1+\frac{0}{2})+1]=0.005$

$k_3=0.1\times[(0+\frac{0.1}{2})-(1+\frac{0.005}{2})+1]=0.004\ 75$

$k_4=0.1\times[(0+0.1)-(1+0.004\ 75)+1]=0.009\ 525$

$y_1=1+\frac{1}{6}[0+2\times(0.005+0.004\ 75)+0.009\ 525]=1.004\ 837\ 50$

仿此计算，所得结果以及与准确解的比较如表 7.3 所示。

表 7.3

| $x_n$ | $y_n$ | $y(x_n)$ | $\lvert y(x_n)-y_n \rvert$ |
|---|---|---|---|
| 0.0 | 1.000 000 00 | 1.000 000 00 | 0 |
| 0.1 | 1.004 837 50 | 1.004 837 42 | $8\times10^{-8}$ |
| 0.2 | 1.018 730 90 | 1.018 730 75 | $1.5\times10^{-7}$ |
| 0.3 | 1.040 818 42 | 1.040 818 22 | $2.0\times10^{-7}$ |
| 0.4 | 1.070 320 29 | 1.070 320 05 | $2.4\times10^{-7}$ |
| 0.5 | 1.106 530 93 | 1.106 530 66 | $2.7\times10^{-7}$ |
| 0.6 | 1.148 811 93 | 1.148 811 64 | $2.9\times10^{-7}$ |
| 0.7 | 1.196 585 62 | 1.196 585 30 | $3.2\times10^{-7}$ |
| 0.8 | 1.249 329 29 | 1.249 328 96 | $3.3\times10^{-7}$ |
| 0.9 | 1.306 569.99 | 1.306 569 66 | $3.3\times10^{-7}$ |
| 1.0 | 1.367 879 77 | 1.367 879 44 | $3.3\times10^{-7}$ |

# §5 线性多步法

上面介绍的台劳级数法和龙格-库塔法都是单步法，它们在计算 $y_{n+1}$ 时只用到 $y_n$ 而没有用到前几步计算得到的信息。为了充分利用这些信息，常使用如下形式的求解公式

$$\sum_{j=0}^{k}\alpha_j y_{n+j} = h\sum_{j=0}^{k}\beta_j f_{n+j} \quad (n = 0,1,2,\cdots,M-k) \tag{7.38}$$

进行计算，通常称为线性多步法。式中 $f_{n+j}=f(x_{n+j},y_{n+j})$，$\alpha_j$、$\beta_j$ $(j=0,1,2,\cdots,k)$ 是与 $n$ 无关的常数，一般 $\alpha_0\neq0,\beta_0\neq0$，不失一般性，可统一取定 $\alpha_k=1$。由式(7.38)可知，在计算 $y_{n+k}$ 值时，需要利用它前面的 $y_{n+k-1},y_{n+k-2},\cdots,y_n$ 的 $k$ 个值，所以称式(7.38)为 $k$ 步法。当 $k=1$ 时为单步法；当 $k>1$ 时为多步法。又因式(7.38)中出现的 $y_{n+j},f_{n+j}$ 都是线性的，所以更确切地说式(7.38)为线性 $k$ 步法。当 $\beta_k=0$ 时，称式(7.38)为显式的；否则称为隐式的。

## 5.1 线性多步法的建立方法

线性多步法有多种建立方法，这里介绍两种方法。

### 5.1.1 利用数值积分的方法

我们对式(7.1)两边积分得

$$\int_{x_{n-1}}^{x_n} y'\mathrm{d}x = \int_{x_{n-1}}^{x_n} f(x,y(x))\mathrm{d}x$$

或

$$y(x_n) = y(x_{n-1}) + \int_{x_{n-1}}^{x_n} y'(x)\mathrm{d}x \tag{7.39}$$

今通过 $x_{n-1},x_{n-2},\cdots,x_{n-k}$ 上的已知值 $y'_{n-1},y'_{n-2},\cdots,y'_{n-k}$ 作拉格朗日插值公式

$$f(x,y(x))=\frac{(x-x_{n-2})\ (x-x_{n-3})\cdots\ (x-x_{n-k})}{(x_{n-1}-x_{n-2})(x_{n-1}-x_{n-3})\cdots(x_{n-1}-x_{n-k})}y'_{n-1}+\cdots+\frac{(x-x_{n-1})\ (x-x_{n-2})\cdots\ (x-x_{n-k+1})}{(x_{n-k}-x_{n-1})(x_{n-k}-x_{n-2})\cdots(x_{n-k}-x_{n-k+1})}y'_{n-k} \tag{7.40}$$

代入式(7.39)后得

$$y_n=y_{n-1}+h(b_1f_{n-1}+b_2f_{n-2}+\cdots+b_kf_{n-k}) \tag{7.41}$$

上式称为显式阿当姆斯公式，其余式为

$$\begin{aligned}R(x_n,h)&=h\int_0^1\frac{u(u+1)\cdots(u+k-1)}{k!}h^ky^{(k+1)}(\tilde{\xi})\mathrm{d}u\\&=h^{k+1}y^{(k+1)}(\xi)\int_0^1\frac{u(u+1)\cdots(u+k-1)}{k!}\mathrm{d}u\\&=Bh^{k+1}y^{(k+1)}(\xi)\qquad \xi\in(x_{n-k},x_{n-1})\end{aligned} \tag{7.42}$$

系数 $b_i(i=1,2,\cdots,k)$ 和误差常数 $B$ 的数值列于表 7.4 中。

表 7.4

| $k$ | $b_1$ | $b_2$ | $b_3$ | $b_4$ | $b_5$ | $b_6$ | $B$ |
|---|---|---|---|---|---|---|---|
| 1 | 1 | | | | | | $\frac{1}{2}$ |
| 2 | $\frac{3}{2}$ | $-\frac{1}{2}$ | | | | | $\frac{5}{12}$ |
| 3 | $\frac{23}{12}$ | $-\frac{16}{12}$ | $\frac{5}{12}$ | | | | $\frac{3}{8}$ |
| 4 | $\frac{55}{24}$ | $-\frac{59}{24}$ | $\frac{37}{24}$ | $-\frac{9}{24}$ | | | $\frac{251}{720}$ |
| 5 | $\frac{1\,901}{720}$ | $-\frac{2\,774}{720}$ | $\frac{2\,616}{720}$ | $-\frac{1\,274}{720}$ | $\frac{251}{720}$ | | $\frac{95}{288}$ |
| 6 | $\frac{4\,277}{1\,440}$ | $-\frac{7\,923}{1\,440}$ | $\frac{9\,982}{1\,440}$ | $-\frac{7\,298}{1\,440}$ | $\frac{2\,877}{1\,440}$ | $-\frac{475}{1\,440}$ | $\frac{10\,987}{60\,480}$ |

按式(7.41)可得

$k=1$： $y_n=y_{n-1}+hf_{n-1}$， $\qquad R(x_n,h)=\frac{1}{2}h^2y''(\xi)$

$k=2$： $y_n=y_{n-1}+\frac{h}{2}(3f_{n-1}-f_{n-2})$， $\qquad R(x_n,h)=\frac{5}{12}h^3y'''(\xi)$

$k=3$： $y_n=y_{n-1}+\frac{h}{12}(23f_{n-1}-16f_{n-2}+5f_{n-3})$， $\qquad R(x_n,h)=\frac{3}{8}h^4y^{(4)}(\xi)$

$k=4$： $y_n=y_{n-1}+\frac{h}{24}(55f_{n-1}-59f_{n-2}+37f_{n-3}-9f_{n-4})$，

$$R(x_n,h)=\frac{251}{720}h^5y^{(5)}(\xi)$$

如果在节点 $x_n,x_{n-1},\cdots,x_{n-k}$ 上建立拉格朗日插值公式代替 $f(x,y(x))$，代入式(7.39)后得隐式阿当姆斯公式

$$y_n = y_{n-1} + h(b_0^* f_n + b_1^* f_{n-1} + \cdots + b_k^* f_{n-k}) \tag{7.43}$$

$$R(x_n, h) = B^* h^{k+2} y^{(k+2)}(\xi), \qquad \xi \in (x_{n-k}, x_n) \tag{7.44}$$

式中的系数值 $b_i^*$ $(i = 0,1,2,\cdots,k)$ 和误差常数 $B^*$ 列于表 7.5 中。

**表 7.5**

| $k$ | $b_0^*$ | $b_1^*$ | $b_2^*$ | $b_3^*$ | $b_4^*$ | $b_5^*$ | $B^*$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | | | | | | $-\frac{1}{2}$ |
| 1 | $\frac{1}{2}$ | $\frac{1}{2}$ | | | | | $-\frac{1}{12}$ |
| 2 | $\frac{5}{12}$ | $\frac{8}{12}$ | $-\frac{1}{12}$ | | | | $-\frac{1}{24}$ |
| 3 | $\frac{9}{24}$ | $\frac{19}{24}$ | $-\frac{5}{24}$ | $\frac{1}{24}$ | | | $-\frac{19}{720}$ |
| 4 | $\frac{251}{720}$ | $\frac{646}{720}$ | $-\frac{264}{720}$ | $\frac{106}{720}$ | $-\frac{19}{720}$ | | $-\frac{3}{160}$ |
| 5 | $\frac{475}{1\,440}$ | $\frac{1\,427}{1\,440}$ | $-\frac{798}{1\,440}$ | $\frac{482}{1\,440}$ | $-\frac{173}{1\,440}$ | $\frac{27}{1\,440}$ | $-\frac{863}{60\,480}$ |

按式(7.43)可得

$k = 0$：$y_n = y_{n-1} + hf_n$，　　$R(x_n, h) = -\frac{1}{2}h^2 y''(\xi)$

$k = 1$：$y_n = y_{n-1} + \frac{h}{2}(f_n + f_{n-1})$，　　$R(x_n, h) = -\frac{1}{12}h^3 y'''(\xi)$

$k = 2$：$y_n = y_{n-1} + \frac{h}{12}(5f_n + 8f_{n-1} - f_{n-2})$，　　$R(x_n, h) = -\frac{1}{24}h^4 y^{(4)}(\xi)$

$k = 3$：$y_n = y_{n-1} + \frac{h}{24}(9f_n + 19f_{n-1} - 5f_{n-2} + f_{n-3})$，　　$R(x_n, h) = -\frac{19}{720}h^5 y^{(5)}(\xi)$

$k = 4$：$y_n = y_{n-1} + \frac{h}{720}(251f_n + 646f_{n-1} - 264f_{n-2} + 106f_{n-3} - 19f_{n-4})$，

$$R(x_n, h) = -\frac{3}{160}h^6 y^{(6)}(\xi)$$

比较表 7.3 和表 7.4 的 $b_i$ 与 $b_i^*$ 的数值表可见，后者比前者绝对值小，因而使用隐式阿当姆斯公式计算时引入的舍入误差也较小。比较它们的余式可知，在同样利用 $k$ 个已知值的情况下，隐式阿当姆斯公式的阶为 $O(h^{k+2})$，而显式阿当姆斯公式的阶为 $O(h^{k+1})$。换言之，为达到同样的误差阶，隐式公式比显式公式少用一个已知值。

### 5.1.2　基于台劳展开式的待定系数法

在下面的台劳展开式

$$y(x_n + h) = y(x_n) + hy'(x_n) + \frac{h^2}{2!}y''(x_n) + \cdots$$

中取前两项得

$$y(x_n + h) \approx y(x_n) + hf(x_n, y(x_n)) \tag{7.45}$$

以等号代替近似号并用 $y_n, y_{n+1}$ 分别代替 $y(x_n), y(x_{n+1})$ 得

$$y_{n+1} = y_n + hf_n \tag{7.46}$$

这是显式线性单步法，即欧拉公式，它正是 $k=1$ 时的显式阿当姆斯公式。如将 $y(x_n+h)$ 和 $y(x_n-h)$ 分别在 $x_n$ 点展开得

$$y(x_n+h)=y(x_n)+hy'(x_n)+\frac{h^2}{2!}y''(x_n)+\frac{h^3}{3!}y'''(x_n)+\cdots$$

$$y(x_n-h)=y(x_n)-hy'(x_n)+\frac{h^2}{2!}y''(x_n)-\frac{h^3}{3!}y'''(x_n)+\cdots$$

将上述两式相减得

$$y(x_n+h)-y(x_n-h)=2hy'(x_n)+\frac{2}{3!}h^3y'''(x_n)+\cdots \tag{7.47}$$

在上式中截取右端一项得

$$y_{n+1}=y_{n-1}+2hf_n \tag{7.48}$$

称为中点公式，其局部截断误差为 $O(h^3)$。

任何一个线性多步法均可使用类似的方法求得。例如对于隐式线性单步法

$$y_{n+1}+\alpha_0y_n=h(\beta_1f_{n+1}+\beta_0f_n) \tag{7.49}$$

只需将 $y_{n+1}$ 与 $y'_{n+1}$ 的展开式

$$y(x_n+h)=y(x_n)+hy'(x_n)+\frac{h^2}{2!}y''(x_n)+\cdots$$

$$f_{n+1}=y'(x_n+h)=y'(x_n)+hy''(x_n)+\frac{h^2}{2!}y'''(x_n)+\cdots$$

代入式(7.49)并合并同类项得

$$c_0y(x_n)+c_1hy'(x_n)+c_2h^2y''(x_n)+\cdots=0 \tag{7.50}$$

其中

$$\begin{cases}c_0=1+\alpha_0\\c_1=1-\beta_1-\beta_0\\c_2=\dfrac{1}{2}-\beta_1\\c_3=\dfrac{1}{6}-\dfrac{1}{2}\beta_1\end{cases} \tag{7.51}$$

为使式(7.49)有尽可能小的局部截断误差，选择

$$\alpha_0=-1,\qquad \beta_0=\beta_1=\frac{1}{2}$$

则 $c_0=c_1=c_2=0$，代入式(7.49)得

$$y_{n+1}=y_n+\frac{h}{2}(f_n+f_{n+1}) \tag{7.52}$$

即为梯形公式，其局部截断误差为 $O(h^3)$。

一般情形下，可按待定系数法来确定一个线性多步法的计算公式。首先建立它的局部截断误差公式

$$R(x,h)=\sum_{j=0}^{k}[\alpha_jy(x+jh)-h\beta_jy'(x+jh)] \tag{7.53}$$

将 $y(x+jh)$ 和 $y'(x+jh)$ 在 $x$ 点的台劳展开式代入上式，合并同类项得

$$R(x,h)=c_0y(x)+c_1hy'(x)+\cdots+c_qh^qy^{(q)}(x)+\cdots \tag{7.54}$$

其中

$$
\begin{cases}
c_0 = \alpha_0 + \alpha_1 + \cdots + \alpha_k \\
c_1 = \alpha_1 + 2\alpha_2 + \cdots + k\alpha_k - (\beta_0 + \beta_1 + \cdots + \beta_k) \\
c_q = \dfrac{1}{q!}(\alpha_1 + 2^q\alpha_2 + \cdots + k^q\alpha_k) - \dfrac{1}{(q-1)!}(\beta_1 + 2^{q-1}\beta_2 + \cdots + k^{q-1}\beta_k) \quad (q = 2,3,\cdots)
\end{cases}
\tag{7.55}
$$

为了衡量线性多步法的公式精度,引入阶的概念。

**定义 7.3**　若在式(7.54)中有

$$c_0 = c_1 = \cdots = c_p = 0 \tag{7.56}$$

而 $c_{p+1} \neq 0$,则称该线性多步法是 $p$ 阶的。

利用式(7.56)便可确定出线性多步法的参数 $\alpha_j, \beta_j (j=0,1,2,\cdots,k)$ 的具体数值。

**例 7.3**　建立显式线性二步法计算公式。

**解**　在本题中,$k=2, \alpha_2=1$,而 $\alpha_0, \beta_0, \beta_1$ 待定。按式(7.55)可列出下面一组关系式

$$c_0 = \alpha_0 + \alpha_1 + 1 = 0$$

$$c_1 = \alpha_1 + 2 - (\beta_0 + \beta_1) = 0$$

$$c_2 = \frac{1}{2!}(\alpha_1 + 4) - \beta_1 = 0$$

设 $\alpha_0 = a$,解得

$$\alpha_1 = -(1+a)$$

$$\beta_1 = \frac{1}{2}(\alpha_1 + 4) = \frac{1}{2}(3-a)$$

$$\beta_0 = -\frac{1}{2}(1+a)$$

而

$$c_3 = \frac{1}{3!}(\alpha_1 + 8) - \frac{1}{2!}\beta_1 = \frac{1}{12}(5+a)$$

由此得显式线性二步法的计算公式为

$$y_{n+2} - (1+a)y_{n+1} + a y_n = \frac{1}{2}h[(3-a)f_{n+1} - (1+a)f_n] \tag{7.57}$$

**例 7.4**　建立具有最高阶数的隐式线性二步法计算公式。

**解**　在本题中,$k=2, \alpha_2=1$,而 $\alpha_0, \beta_0, \beta_1, \beta_2$ 待定。按式(7.55)建立关系式

$$c_0 = \alpha_0 + \alpha_1 + 1 = 0$$

$$c_1 = \alpha_1 + 2 - (\beta_0 + \beta_1 + \beta_2) = 0$$

$$c_2 = \frac{1}{2!}(\alpha_1 + 4) - (\beta_1 + 2\beta_2) = 0$$

$$c_3 = \frac{1}{3!}(\alpha_1 + 8) - \frac{1}{2!}(\beta_1 + 4\beta_2)$$

设 $\alpha_0 = a$,解得

$$\alpha_1 = -(1+a)$$

$$\beta_0 = -\frac{1}{12}(1+5a)$$

$$\beta_1 = \frac{2}{3}(1-a)$$

$$\beta_2 = \frac{1}{12}(5+a)$$

而
$$c_4=\frac{1}{4!}(\alpha_1+16)-\frac{1}{3!}(\beta_1+8\beta_2)=-\frac{1}{4!}(1+a)$$
$$c_5=\frac{1}{5!}(\alpha_1+32)-\frac{1}{4!}(\beta_1+16\beta_2)=-\frac{1}{3\times 5!}(17+13a)$$

若 $a\neq -1$，则 $c_4\neq 0$，方法是三阶的；若 $a=-1$，则 $c_4=0,c_5\neq 0$，方法是四阶的。由此得到阶数最高的隐式线性二步法为

$$y_{n+2}=y_n+\frac{h}{3}[f_{n+2}+4f_{n+1}+f_n] \tag{7.58}$$

这正是辛卜生求积公式。当 $a=0$ 时，相应的线性二步法为

$$y_{n+2}=y_{n+1}+\frac{h}{12}(5f_{n+2}+8f_{n+1}-f_n) \tag{7.59}$$

它就是 $k=2$ 的隐式阿当姆斯公式。当 $a=-5$ 时，可得显式二步法计算公式。

形如

$$\sum_{j=0}^{k}\alpha_j y_{n+j}=h\beta_k f_{n+k} \tag{7.60}$$

的线性 $k$ 步法称为吉尔(Gear)方法，按照式(7.60)的形式和式(7.55)可构造 $k$ 从 1 到 6 的 $k$ 阶 $k$ 步法。例如，对于二阶二步吉尔方法，在式(7.60)中 $k=2$，取 $\alpha_2=1$，由式(7.55)，令

$$\begin{cases}c_0=\alpha_0+\alpha_1+1=0\\ c_1=\alpha_1+2-\beta_2=0\\ c_2=\dfrac{1}{2}(\alpha_1+4)-2\beta_2=0\end{cases}$$

解得 $\alpha_1=-\frac{4}{3}$， $\alpha_0=\frac{1}{3}$， $\beta_2=\frac{2}{3}$。因

$$c_3=\frac{1}{3!}(\alpha_1+8\alpha_2)-\frac{1}{2}(4\beta_2)=-\frac{2}{9}\neq 0$$

故得二阶二步吉尔方法为

$$y_{n+2}-\frac{4}{3}y_{n+1}+\frac{1}{3}y_n=\frac{2}{3}hf_{n+2}$$

$k$ 从 1 到 6 的吉尔方法的系数见表 7.6。

**表 7.6**

| $k$ | $\alpha_6$ | $\alpha_5$ | $\alpha_4$ | $\alpha_3$ | $\alpha_2$ | $\alpha_1$ | $\alpha_0$ | $\beta_k$ |
|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | 1 | $-1$ | 1 |
| 2 | | | | | 1 | $-\frac{4}{3}$ | $\frac{1}{3}$ | $\frac{2}{3}$ |
| 3 | | | | 1 | $-\frac{18}{11}$ | $\frac{9}{11}$ | $-\frac{2}{11}$ | $\frac{6}{11}$ |
| 4 | | | 1 | $-\frac{48}{25}$ | $\frac{36}{25}$ | $-\frac{16}{25}$ | $\frac{3}{25}$ | $\frac{12}{25}$ |
| 5 | | 1 | $-\frac{300}{137}$ | $\frac{300}{137}$ | $-\frac{200}{137}$ | $\frac{75}{137}$ | $-\frac{12}{137}$ | $\frac{60}{137}$ |
| 6 | 1 | $-\frac{360}{147}$ | $\frac{450}{147}$ | $-\frac{400}{147}$ | $\frac{255}{147}$ | $-\frac{72}{147}$ | $\frac{10}{147}$ | $\frac{60}{147}$ |

## 5.2　起始值的求取方法

在采用 $p$ 阶线性 $k(>1)$ 步法求解初值问题(7.1)时，必须附加 $y_1, y_2, \cdots, y_{k-1}$ 个起始值才能按线性 $k$ 步法逐次递推以下的数值解。这些起始值的精度至少不低于该方法的局部截断误差 $O(h^{p+1})$。下面介绍几种计算起始值(亦称表头)的具体方法。

### 5.2.1　台劳级数法

本法的具体内容已在本章§2中介绍过，它本身也是一种求取数值解的方法，因其计算工作量较大，通常只用它求取起始几点的数值解。

另外，在使用台劳级数法作表头时，还可以得到选择步长 $h$ 的信息。假定要求局部截断误差不超过 $\varepsilon$，那么，当 $h$ 满足

$$\frac{1}{(p+1)!}h^{p+1}\left|y^{(p+1)}(x_n)\right| \leqslant \varepsilon$$

及

$$\frac{1}{p!}h^{p}\left|y^{(p)}(x_n)\right| > \varepsilon$$

时，认为是适当的。

### 5.2.2　含有高阶导数的隐式线性单步法

我们从 $(x_n, y_n)$ 点出发，使用如下一组导数值

$$\begin{cases} y'_n, y''_n, y'''_n, \cdots, y_n^{(l)} \\ y'_{n+1}, y''_{n+1}, y'''_{n+1}, \cdots, y_{n+1}^{(l)} \end{cases} \tag{7.61}$$

按以下方法进行组合

$$\alpha_1 y_{n+1} + \alpha_0 y_n = h(\beta_{11} y'_{n+1} + \beta_{10} y'_n) + h^2(\beta_{21} y''_{n+1} + \beta_{20} y''_n) + \cdots + h^l(\beta_{l1} y_{n+1}^{(l)} + \beta_{l0} y_n^{(l)}) \tag{7.62}$$

对于给定的 $l$ 值，可仿线性多步法求解公式的建立方法选取适当的 $\alpha_j, \beta_{sj}$ $(j=0,1; s=1,2,\cdots,l)$，使式(7.62)具有尽可能小的局部截断误差，以下引用几个这样的公式供选用。

$$l=2: \begin{cases} y_{n+1} - y_n = \dfrac{1}{2}h(y'_{n+1} + y'_n) - \dfrac{1}{12}h^2(y''_{n+1} - y''_n) \\ R(x_n, h) = \dfrac{1}{720}h^5 y^{(5)}(x_n) + O(h^6) \end{cases} \tag{7.63}$$

$$l=3: \begin{cases} y_{n+1} - y_n = \dfrac{1}{2}h(y'_{n+1} + y'_n) - \dfrac{1}{10}h^2(y''_{n+1} - y''_n) + \dfrac{1}{120}h^3(y'''_{n+1} + y'''_n) \\ R(x_n, h) = -\dfrac{1}{100\ 800}h^7 y^{(7)}(x_n) + O(h^8) \end{cases} \tag{7.64}$$

$$l=4: \begin{cases} y_{n+1} - y_n = \dfrac{1}{2}h(y'_{n+1} + y'_n) - \dfrac{3}{28}h^2(y''_{n+1} - y''_n) + \dfrac{1}{84}h^3(y'''_{n+1} + y'''_n) - \\ \qquad \dfrac{1}{1\ 680}h^4(y_{n+1}^{(4)} - y_n^{(4)}) \\ R(x_n, h) = \dfrac{1}{25\ 401\ 600}h^9 y^{(9)}(x_n) + O(h^{10}) \end{cases} \tag{7.65}$$

式中的高阶导数可按式(7.6)计算。以上这些公式都是隐式公式，求解时需用迭代法(参见本节5.3)。

## 5.3 线性多步法的数值求解方法

### 5.3.1 显式公式的求解方法

线性多步法的显式公式求解前，必须计算好表头，然后便可按显式公式逐次递推出以下各数值解。

**例 7.5** 求初值问题

$$\begin{cases} y' = x - y + 1 \\ y(0) = 1 \end{cases}$$

的数值解。

**解** 精确解为 $y=e^{-x}+x$。今取步长 $h=0.1$，用显式四阶阿当姆斯公式求解之。

$$y_{n+1} = y_n + \frac{h}{24}[55f(x_n, y_n) - 59f(x_{n-1}, y_{n-1}) + 37f(x_{n-1}, y_{n-1}) - 9f(x_{n-3}, y_{n-3})]$$

$$(n = 3,4,\cdots,) \tag{7.66}$$

因为 $f(x,y)=x-y+1$，$h=0.1$，$x_n=0.1n$，代入上式得

$$y_{n+1}=\frac{1}{24}(18.5y_n+5.9y_{n-1}-3.7y_{n-2}+0.9y_{n-3}+0.24n+2.52) \qquad (n=3,4,\cdots,) \tag{7.67}$$

如表头用精确解计算，则得

$$\begin{aligned} &x_0=0.0, \quad && y_0=1 \\ &x_1=0.1, && y_1=e^{-0.1}+0.1=1.004\,837\,418 \\ &x_2=0.2, && y_2=e^{-0.2}+0.2=1.018\,730\,753 \\ &x_3=0.3, && y_3=e^{-0.3}+0.3=1.040\,818\,221 \end{aligned}$$

令 $n=3$，按式(7.67)计算 $x_4=0.4$ 的 $y_4$ 值

$$y_4 = \frac{1}{24}(18.5y_3 + 5.9y_2 - 3.7y_1 + 0.9y_0 + 0.24 \times 3 + 2.52) = 1.070\,322\,92$$

令 $n=4$，继续递推得

$$y_5 = \frac{1}{24}(18.5y_4 + 5.9y_3 - 3.7y_2 + 0.9y_1 + 0.24 \times 4 + 2.52) = 1.106\,535\,476$$

…

### 5.3.2 隐式公式的求解方法

由前知，在显式的线性多步法中，$y_{n+k}$值是按前面已知的 $k$ 个值进行推算的；而在隐式的线性多步法中，情形就不同了。这是因为在式(7.38)的右边含有 $f(x_{n+k}, y_{n+k})$项，而 $y_{n+k}$恰是待求的未知量，这时式(7.38)是变量 $y_{n+k}$的隐式方程，可表为

$$y_{n+k} = h\beta_k f(x_{n+k}, y_{n+k}) + g_n \tag{7.68}$$

式中，$g_n$ 为已知的数值。求解上式可采用迭代法

$$y_{n+k}^{(r+1)} = h\beta_k f^{(r)}(x_{n+k}, y_{n+k}) + g_n \tag{7.69}$$

式中，$y_{n+k}^{(0)}$为初值。由迭代解法知，当下列条件

$$\left| h\beta_k \frac{\partial f}{\partial y} \right| < 1 \tag{7.70}$$

或

$$hL < \frac{1}{|\beta_k|} \qquad \left( \left| \frac{\partial f}{\partial y} \right| < L \right) \tag{7.71}$$

满足时，上述迭代过程是收敛的。

在选用隐式阿当姆斯公式时，$\beta_k = b_0^*$，相应于不同的 $k$ 值有

$$\begin{cases} k=1, & |hL| < 2 \\ k=2, & |hL| < \dfrac{12}{5} \\ k=3, & |hL| < \dfrac{8}{3} \\ k=4, & |hL| < \dfrac{720}{251} \\ k=5, & |hL| < \dfrac{1\,440}{475} \end{cases} \tag{7.72}$$

因此只要适当取定 $h$ 值，总能使式(7.72)成立，从而保证迭代过程收敛。一般迭代次数不宜过多，否则会增加计算量。迭代次数以 2～3 次为宜，如过多，则应减小 $h$ 值。

(1) 隐式公式的迭代解法

对于隐式公式(7.69)，初值可取 $y_{n+k}^{(0)} = g_n$，例如对隐式二阶阿当姆斯公式

$$y_{n+1} = y_n + \frac{h}{2}[f(x_n, y_n) + f(x_{n+1}, y_{n+1})] \tag{7.73}$$

可取

$$y_{n+1}^{(0)} = y_n + \frac{h}{2} f(x_n, y_n) \tag{7.74}$$

再利用隐式公式进行迭代

$$y_{n+1}^{(r+1)} = y_n + \frac{h}{2}[f(x_n, y_n) + f(x_{n+1}, y_{n+1}^{(r)})]$$

直到满足精度要求为止。

**例 7.6**　用式(7.73)求解

$$y' = y^2,\ y(0) = 1$$

**解**　计算公式如下

$$\begin{cases} y_{n+1}^{(0)} = y_n + h y_n^2 \\ y_{n+1}^{(r+1)} = y_n + \dfrac{h}{2}[y_n^2 + (y_{n+1}^{(r)})^2] \end{cases} \tag{7.75}$$

取步长 $h=0.1$，则有

$$y_1^{(0)} = 1 + 0.1 \times 1^2 = 1.1$$

$$y_1^{(1)} = 1 + \frac{0.1}{2}(1^2 + 1.1^2) = 1.110\,5$$

$$y_1^{(2)} = 1 + \frac{0.1}{2}(1^2 + 1.110\,5^2) = 1.111\,7$$

$$y_1^{(3)} = 1 + \frac{0.1}{2}(1^2 + 1.111\,7^2) = 1.111\,8$$

$$y_1^{(4)} = 1 + \frac{0.1}{2}(1^2 + 1.111\,8^2) = 1.111\,8$$

取 $x_1 = 0.1$，$y_1 = 1.111\,8$，同法类推可得

$$x_2 = 0.2,\ y_2 = 1.252\,1$$
$$x_3 = 0.3,\ y_3 = 1.434\,5$$

$$x_4 = 0.4,\ y_4 = 1.678\,2$$

(2) 预测-校正法

隐式公式的迭代解法要多次计算 $f(x,y)$ 的函数值，这样要花费较大的代价。因此可以考虑用一个显式公式提供 $y_{n+1}$ 的良好估计(称为预测)，代入隐式公式中得到 $y_{n+1}$ 的校正值(称为校正)，这就组成了预测-校正法。虽然前述的迭代解法也可以看成是不断进行校正的过程，但它的迭代次数是由 $|y_{n+1}^{(r+1)} - y_{n+1}^{(r)}| < \varepsilon$ 来控制的，这种方法的局部截断误差和数值稳定性主要取决于隐式公式的特性。而预测-校正法则是事先规定了预测与校正的次数，当然仍然要求用这种方法求得的数值解满足精度要求。

计算 $y_{n+1}$ 的预测-校正法步骤如下。

① 利用预测公式计算 $y_{n+1}$ 的初始近似值 $y_{n+1}^{(0)}$，令 $i=0$。

② 计算右端函数 $f_{n+1}^{(i)} = f(x_{n+1}, y_{n+1}^{(i)})$。

③ 将 $f_{n+1}^{(i)}$ 代入校正公式，计算出新的近似值 $y_{n+1}^{(i+1)}$。

④ 校正次数 $\leqslant m$?，如成立，$i$ 增 1 后转②；否则令 $y_{n+1} = y_{n+1}^{(i+1)}$。

这里出现了三个互相独立的过程：预测步骤①，我们称之为 P；计算步骤②，称为 E；校正步骤③，称为 C。步骤④由 $m$ 次迭代组成。这种方法称为 $\mathrm{P(EC)}^m$ 法。其中 $\mathrm{(EC)}^m$ 表示 E，C 过程重复计算 $m$ 次。许多事实表明，最后再作一次函数计算为下一步提供一个更精确的 $f$ 值是值得的，这种方法表示为 $\mathrm{P(EC)}^m\mathrm{E}$ 法。大多数使用的预测-校正法在每步上最多进行两次函数值计算，因此下面的一些计算方案 PEC，PECE，$\mathrm{P(EC)}^2$ 用得比较普遍。在预测-校正法中，预测公式的局部截断误差的阶较校正公式的局部截断误差的阶略低一些，一般取低一阶或相等，此时经预测、校正一次后的误差阶与校正公式相同。特别当预测公式与校正公式的阶数相同时，可以得到预测公式或校正公式的局部截断误差主部的事后误差估计式。

常用的预测-校正公式有(* 标记预测公式)如下几种。

① 改进的欧拉公式

$$\begin{cases} y_{n+1}^* = y_n + hf(x_n, y_n) \\ y_{n+1} = y_n + \dfrac{h}{2}[f(x_n, y_n) + f(x_{n+1}, y_{n+1})] \\ R^*(x_n, h) = \dfrac{h}{2} y''(\xi_1) \\ R(x_n, h) = -\dfrac{1}{12} h^2 y'''(\xi_2) \end{cases} \tag{7.76}$$

② 三点米尔纳(Milne)法

$$\begin{cases} y_{n+1}^* = y_{n-3} + \dfrac{h}{3}[8f(x_n, y_n) - 4f(x_{n-1}, y_{n-1}) + 8f(x_{n-2}, y_{n-2})] \\ y_{n+1} = y_{n-1} + \dfrac{h}{3}[f(x_{n+1}, y_{n+1}) + 4f(x_n, y_n) + f(x_{n-1}, y_{n-1})] \\ R^*(x_n, h) = \dfrac{28}{90} h^4 y^{(5)}(\xi_1) \\ R(x_n, y) = -\dfrac{1}{90} h^4 y^{(5)}(\xi_2) \end{cases} \tag{7.77}$$

③ 阿当姆斯四阶方法

$$\begin{cases} y_{n+1}^{*} = y_n + \frac{h}{24}[55f(x_n, y_n) - 59f(x_{n-1}, y_{n-1}) + 37f(x_{n-2}, y_{n-2}) - 9f(x_{n-3}, y_{n-3})] \\ y_{n+1} = y_n + \frac{h}{24}[9f(x_{n+1}, y_{n+1}) + 19f(x_n, y_n) - 5f(x_{n-1}, y_{n-1}) + f(x_{n-2}, y_{n-2})] \\ R^{*}(x_n, h) = \frac{251}{720}h^5 y^{(5)}(\xi_1) \\ R(x_n, h) = -\frac{19}{720}h^5 y^{(5)}(\xi_2) \end{cases} \tag{7.78}$$

④五点米尔纳(Milne)公式

$$\begin{cases} y_{n+1}^{*} = y_{n-5} + \frac{h}{10}[33f(x_n, y_n) - 42f(x_{n-1}, y_{n-1}) + 78f(x_{n-2}, y_{n-2}) - \\ \qquad 42f(x_{n-3}, y_{n-3}) + 33f(x_{n-4}, y_{n-4})] \\ y_{n+1} = y_{n-3} + \frac{h}{45}[14f(x_{n+1}, y_{n+1}) + 64f(x_n, y_n) + 24f(x_{n-1}, y_{n-1}) + \\ \qquad 64f(x_{n-2}, y_{n-2}) + 14f(x_{n-3}, y_{n-3})] \\ R^{*}(x_n, h) = \frac{41}{140}h^6 y^{(7)}(\xi_1) \\ R(x_n, h) = -\frac{8}{945}h^6 y^{(7)}(\xi_2) \end{cases} \tag{7.79}$$

**例 7.7**　用阿当姆斯四阶预测-校正法解初值问题

$$y' = x - y + 1,\ y(0) = 1$$

**解**　取 $h=0.1$，开始几点值用龙格-库塔法计算(见表 7.3)得

$x_0=0.0$，　$y_0=1$，　$f_0=0$
$x_1=0.1$，　$y_1=1.004\ 837\ 5$，　$f_1=0.095\ 162\ 5$
$x_2=0.2$，　$y_2=1.018\ 730\ 90$，　$f_2=0.181\ 269\ 1$
$x_3=0.3$，　$y_3=1.040\ 818\ 42$，　$f_3=0.259\ 181\ 58$

以下每步按预测、计算、校正计算。

预测　$y_{n+1}^{*} = y_n + \frac{h}{24}(55f_n - 59f_{n-1} + 37f_{n-2} - 9f_{n-3})$

计算　$E_{n+1} = f(x_{n+1}, y_{n+1}^{*})$

校正　$y_{n+1} = y_n + \frac{h}{24}(9E_{n+1} + 19f_n - 5f_{n-1} + f_{n-2})$

$x_4=0.4$，　$y_4^{*} = y_3 + \frac{h}{24}(55f_3 - 59f_2 + 37f_1 - 9f_0) = 1.070\ 323\ 097$

$E_4 = 0.4 - 1.070\ 323\ 097 + 1 = 0.329\ 676\ 903$

$y_4 = y_3 + \frac{h}{24}(9E_4 + 19f_3 - 5f_2 + f_1) = 1.070\ 319\ 916$

继续推算可得

$x_5=0.5$，　$y_5=1.106\ 530\ 27$
$x_6=0.6$，　$y_6=1.148\ 811\ 03$
$x_7=0.7$，　$y_7=1.196\ 584\ 53$
…

为了在预测-校正法中避免迭代以减少计算工作量，并且保证数值解的精度，可采用事后估计

误差对预测值或校正值进行修正。为便于研制修正公式，选取预测公式与校正公式具有相同阶的局部截断误差。以阿当姆斯四阶预测-校正法为例，预测公式和校正公式的局部截断误差分别为

$$y(x_{n+1})-P_{n+1}\approx\frac{251}{720}h^5y^{(5)}(\xi_1) \tag{7.80}$$

$$y(x_{n+1})-C_{n+1}\approx-\frac{19}{720}h^5y^{(5)}(\xi_2) \tag{7.81}$$

式中，$P_{n+1}$为预测值，$C_{n+1}$为校正值。假定 $y^{(5)}(\xi_1)\approx y^{(5)}(\xi_2)$，经过简单的运算，可得事后误差估计

$$y(x_{n+1})-P_{n+1}\approx-\frac{251}{270}(P_{n+1}-C_{n+1}) \tag{7.82}$$

$$y(x_{n+1})-C_{n+1}\approx-\frac{19}{270}(C_{n+1}-P_{n+1}) \tag{7.83}$$

这样，如果在预测值 $P_{n+1}$上加上局部截断误差的估计值$\frac{251}{270}(C_n-P_n)$（因尚无 $C_{n+1}$，$C_{n+1}-P_{n+1}$用 $C_n-P_n$ 代替），在校正值 $C_{n+1}$上加上$-\frac{19}{270}(C_{n+1}-P_{n+1})$，结果的精度将得到改善。于是得到预测-修正-校正-修正法如下。

$$\begin{array}{ll}
\text{预测} & P_{n+1}=y_n+\frac{h}{24}(55f_n-59f_{n-1}+37f_{n-2}-9f_{n-3}) \\
\text{修正} & m_{n+1}=P_{n+1}+\frac{251}{270}(C_n-P_n) \\
\text{计算} & E_{n+1}=f(x_{n+1},m_{n+1}) \\
\text{校正} & C_{n+1}=y_n+\frac{h}{24}(9E_{n+1}+19f_n-5f_{n-1}+f_{n-2}) \\
\text{修正} & y_{n+1}=C_{n+1}-\frac{19}{270}(C_{n+1}-P_{n+1}) \\
\text{计算} & f_{n+1}=f(x_{n+1},y_{n+1})
\end{array} \tag{7.84}$$

（式中开始时的 $C_n-P_n$ 可令其为 0。）

上述方法不仅改善了计算结果，而且 $C_n-P_n$ 还可用来估计每一步的局部截断误差和调整步长。在舍入误差较小的情况下，如果$|C_n-P_n|$很小，则可适当放大步长，反之亦可缩小步长。如果$|C_n-P_n|$出现突然变化，说明计算中有了问题，可以帮助我们去检查计算中的问题。在改变步长时，应从该点起按新步长重建表头，再继续计算。

这类公式还有以下几种。

① 米尔纳(Milne)预测-修正-校正-修正公式

$$\begin{cases}
\text{预测} & P_{n+1}=y_{n-3}+\frac{h}{3}(8f_n-4f_{n-1}+8f_{n-2}) \\
\text{修正} & m_{n+1}=P_{n+1}+\frac{28}{29}(C_n-P_n) \\
\text{计算} & E_{n+1}=f(x_{n+1},m_{n+1}) \\
\text{校正} & C_{n+1}=y_{n-1}+\frac{h}{3}(E_{n+1}+4f_n+f_{n-1}) \\
\text{修正} & y_{n+1}=C_{n+1}-\frac{1}{29}(C_{n+1}-P_{n+1}) \\
\text{计算} & f_{n+1}=f(x_{n+1},y_{n+1})
\end{cases} \tag{7.85}$$

② 哈明预测-修正-校正-修正公式

$$
\begin{cases}
\text{预测} & P_{n+1}=y_{n-3}+\dfrac{4h}{3}(2f_n-f_{n-1}+2f_{n-2}) \\
\text{修正} & m_{n+1}=P_{n+1}+\dfrac{112}{121}(C_n-P_n) \\
\text{计算} & E_{n+1}=f(x_{n+1},m_{n+1}) \\
\text{校正} & C_{n+1}=\dfrac{9}{8}y_n-\dfrac{1}{8}y_{n-2}+\dfrac{3h}{8}(E_{n+1}+2f_n-f_{n-1}) \\
\text{修正} & y_{n+1}=C_{n+1}-\dfrac{9}{121}(C_{n+1}-P_{n+1}) \\
\text{计算} & f_{n+1}=f(x_{n+1},y_{n+1})
\end{cases}
\tag{7.86}
$$

## 5.4　线性多步法的步长改变方法

在线性 $k$ 步法中，如需要改变步长时，应注意到在新的步长下的 $k$ 个起始值是否已经准备好，即要重新做新步长下的表头值。本章前面介绍的几种起始值的算法均可用作中途变更步长之用。即以当前的数值为初值，取新的步长计算出 $k-1$ 个新的起始值后继续按线性 $k$ 步法递推以下的数值解。下面再介绍两种变步长方法供选用。

### 5.4.1　插值法

以线性四步法为例，设原步长为 $h$，若要求用 $2h$ 为步长，由 $x_{n+3}$ 点出发计算新 $y_{n+4}$ 时，则需要使用 $x_{n+3},x_{n+1},x_{n-1},x_{n-3}$ 点上的有关数据，因它们已经算出，可直接取用。若要求用 $h/2$ 为步长计算新值 $y_{n+4}$ 时，则需要使用 $x_{n+3},x_{n+2.5},x_{n+2},x_{n+1.5}$ 点上的 $y$ 与 $f$ 值。其中 $x_{n+2.5}$，$x_{n+1.5}$ 点上的 $y$ 与 $f$ 为未知值，可以取已知的数值解建立插值公式算出所需的 $y$ 值，代人 $f(x,y)$ 求出相应的 $f$ 值。在建立插值公式时，要求插值公式的余式与线性多步法的局部截断误差有同样的阶数。

### 5.4.2　线性组合法

我们亦可仿照线性多步法的建立方法中基于台劳展开式的待定系数法，按照方法的阶数，用已知数据的线性组合来代替插值公式计算所需的 $y$ 值或 $f$ 值。以显式三步阿当姆斯公式为例，设原步长为 $h$，当前点为 $x_{n+2}$，今欲以新步长 $h/2$ 计算新值 $y_{n+2.5}$，其所需的 $f_{n+1.5}$ 可用 $f_{n+2}$、$f_{n+1}$、$f_n$ 的线性组合表示为

$$f_{n+1.5}=\frac{3}{8}f_{n+2}+\frac{3}{4}f_{n+1}-\frac{1}{8}f_n+O(h^4) \tag{7.87}$$

若视当前点 $x_{n+2}$ 为 $t$ 时，则上式可表示为

$$y'(t-\frac{h}{2})=\frac{3}{8}y'(t)+\frac{3}{4}y'(t-h)-\frac{1}{8}y'(t-2h)+O(h^4) \tag{7.88}$$

这样便得到以 $h/2$ 为间距的三个数据 $y'(t-h)$、$y'(t-h/2)$、$y'(t)$，以下即可以 $h/2$ 为步长继续推算数值解。

另有一种组合方法如下，首先利用线性组合公式

$$y(t+\frac{h}{2})=y(t)+\frac{h}{24}[17y'(t)-7y'(t-h)+2y'(t-2h)]+O(h^4) \tag{7.89}$$

计算出新点$(x_{n+2.5},y_{n+2.5})$，用$(x_{n+3},y_{n+3})$表示，再视当前点 $x_{n+3}$ 为 $t$，利用线性组合公式

$$y(t+\frac{h}{2}) = y(t) + \frac{h}{72}[64y'(t) - 33y'(t-\frac{h}{2}) + 5y'(t-\frac{3}{2}h)] + O(h^4) \quad (7.90)$$

计算出 $x_{n+4}$ 点上的 $y_{n+4}$ 值，代入 $f(x,y)$ 可以计算出相应的 $f_{n+3}$、$f_{n+4}$，这时 $x_{n+2}$、$x_{n+3}$、$x_{n+4}$ 便是间距为 $h/2$ 的三个节点，以下即可以 $h/2$ 为步长推算其数值解。

上述两种组合法所得的结果是不相同的，方法是不等价的。它们在整个稳定性的性质上也可能是不同的，对于变步长多步法的稳定性问题还有待进一步的探索和研究。

以上变步长方法可组合到预测-校正法中，可使局部截断误差在小于给定 $\varepsilon$ 值的前提下达到计算量最省的目的。

## §6 单步法的收敛性、相容性与稳定性

显式单步法的一般形式可以统一表为

$$y_{n+1} = y_n + h\varphi(x_n, y_n, h) \qquad (n = 0,1,2,\cdots,M-1) \quad (7.91)$$

式中，函数 $\varphi(x,y,h)$ 与函数 $f$ 有关，称为增量函数。

### 6.1 单步法的收敛性

我们用差分方程近似微分方程来求取其初值问题的数值解，那么此数值解是否收敛到微分方程的精确解的问题就是收敛性问题，具体定义如下。

**定义 7.4** 对于满足定理 7.1 的初值问题(7.1)，如果由某个单步方法式(7.91)产生的近似解，对于任一固定的点 $x_i = x_0 + ih$，当 $h \to 0$（同时 $i \to \infty$）时，有

$$\lim_{h\to 0} y_i = y(x_i) \quad (7.92)$$

则称该数值方法是收敛的；否则称为不收敛。

此定义对单步隐式法与多步法同样适用。从上述定义可知，若数值方法式(7.91)收敛，则其局部截断误差趋于 0。根据定义，数值方法的收敛性需根据该方法的整体截断误差来判定。

**定理 7.3** 若初值问题(7.1)使用单步法式(7.91)时的局部截断误差为 $O(h^{p+1})$ ($p \geqslant 1$)，且 $\varphi(x,y,h)$ 对 $y$ 满足李普希兹条件，即存在 $k>0$，下式

$$|\varphi(x,y_1,h) - \varphi(x,y_2,h)| \leqslant k|y_1 - y_2| \quad (7.93)$$

对一切 $y_1, y_2$ 成立，则单步法式(7.91)收敛，且单步法的整体截断误差为

$$\varepsilon_{n+1} = O(h^p) \quad (7.94)$$

**证** 我们把单步法式(7.91)在点 $x_{n+1}$ 处的整体截断误差表为

$$\begin{aligned}\varepsilon_{n+1} &= y(x_{n+1}) - y_{n+1} = y(x_{n+1}) - y(x_n) - h\varphi(x_n, y(x_n), h) + \\ &\quad y(x_n) - [y_n + h\varphi(x_n, y_n, h)] + h\varphi(x_n, y(x_n), h) \\ &= \{y(x_{n+1}) - [y(x_n) + h\varphi(x_n, y(x_n), h)]\} + [y(x_n) - y_n] + \\ &\quad h[\varphi(x_n, y(x_n), h) - \varphi(x_n, y_n, h)]\end{aligned}$$

则有

$$|\varepsilon_{n+1} \leqslant |R_{n+1}| + |y(x_n) - y_n| + hk|y(x_n) - y_n|$$

$$|\varepsilon_{n+1}| \leqslant |R_{n+1}| + \varepsilon_n(1 + hk) \quad (7.95)$$

其中 $R_{n+1} = O(h^{p+1})$。反复利用式(7.95)得

$$|\varepsilon_{n+1}| \leqslant |R_{n+1}| + |R_n|(1+hk) + |\varepsilon_{n-1}|(1+hk)^2$$

$$\leqslant \cdots \leqslant \sum_{k=0}^{n} |R_{n+1-k}| (1+hk)^k + |\varepsilon_0| (1+hk)^{n+1}$$

记

$$R = \max_{1 \leqslant n < M} |R_n|$$

并注意到 $\varepsilon_0 = 0$，就有

$$\begin{aligned}
|\varepsilon_{n+1}| \leqslant R\sum_{k=0}^{n}(1+hk)^k &= \frac{R}{hk}[(1+hk)^{n+1}-1] \\
&= \frac{R}{hk}\{[(1+hk)^{\frac{1}{hk}}]^{(n+1)hk}-1\} \\
&\leqslant \frac{R}{hk}[e^{(n+1)hk}-1] \\
&\leqslant \frac{R}{hk}[e^{Mhk}-1] \\
&= \frac{R}{hk}[e^{(b-a)k}-1] \\
&= \frac{1}{hk}|O(h^{p+1})|[e^{(b-a)k}-1] = |O(h^p)|
\end{aligned} \tag{7.96}$$

对上式取极限得

$$\lim_{h\to 0} |\varepsilon_{n+1}| = 0$$

从而证得，对区间$[a,b]$中的任意一点 $x_{n+1}$，当 $h\to 0$ 时，一致地有

$$\lim_{h\to 0} y_{n+1} = y(x_{n+1})$$

亦即数值解 $y_{n+1}$一致收敛于初值问题(7.1)的理论解 $y(x_{n+1})$。

从上面的推证过程可知，数值解的整体截断误差与起始值的误差有关，还与参数 $k$ 值及局部截断误差 $O(h^{p+1})$中的参数有关，由于这些参数很难估计，而且在推导中一再放大误差上限，这样的估计很保守，远远大于实际的误差，所以估计式(7.96)难以实际使用。但是从这个估计式中可以得到下面有用的结论：整体截断误差的阶数比局部截断误差的阶数低一次。因此，局部截断误差的阶数大小可以表征单步法式(7.91)的精度高低；它同时也表明可以从提高局部截断误差的阶数入手去构造精度较高的数值解法。但若函数 $y(x)$本身具有连续导数的阶数不高，则宜采用低阶方法。

### 6.2　单步法的相容性

在单步法式(7.91)中，如果把 $y_n$ 和 $y_{n+1}$分别换成 $y(x)$与 $y(x+h)$，则得到关于 $y(x)$的一个近似方程

$$\frac{y(x+h)-y(x)}{h} \approx \varphi(x,y(x),h) \tag{7.97}$$

差分方程(7.91)的解 $y_1, y_2, \cdots, y_M$ 能否作为初值问题(7.1)的近似解，应取决于当 $h\to 0$ 时，近似方程(7.97)的极限状态能否成为微分方程(7.1)。

由于

$$\lim_{h\to 0}\frac{y(x+h)-y(x)}{h} = y'(x)$$

因此，要使近似方程(7.97)的极限状态成为微分方程，只需

$$\lim_{h\to0}\varphi(x,y(x),h)=f(x,y(x)) \tag{7.98}$$

成立。总假定 $\varphi(x,y(x),h)$ 是连续函数，因而上式可表示为

$$\varphi(x,y(x),0)=f(x,y(x)) \tag{7.99}$$

**定义 7.5** 单步法式(7.91)称为与微分方程(7.1)相容，如果条件式(7.99)成立；并称条件(7.99)为相容条件。

相容性可理解为，单步法式(7.91)的解确实是微分方程(7.1)的近似解，而不是其他方程的近似解。注意到欧拉公式中

$$\varphi(x,y,h)=f(x,y) \tag{7.100}$$

以及后退的欧拉公式中

$$\varphi(x,y,h)=f(x+h,y(x+h)) \tag{7.101}$$

它们均有 $\varphi(x,y,0)=f(x,y)$ 成立，即它们都是相容的方法。

对于龙格-库塔公式来说，其一般公式可以表示为

$$y_{n+1}=y_n+c_1k_1+c_2k_2+\cdots+c_\nu k_\nu \tag{7.102}$$

式中，$k_i=hf(x_n+\lambda_ih,y_n+\sum\limits_{j=1}^{i-1}\mu_jk_j)$。所以有

$$h\varphi(x,y,h)=c_1k_1+c_2k_2+\cdots+c_\nu k_\nu \tag{7.103}$$

因 $\lim\limits_{h\to0}k_i=hf(x,y)$，对式(7.103)两边取极限($h\to0$)推得

$$\varphi(x,y,0)=\sum_{i=1}^{\nu}c_if(x,y)=(\sum_{i=1}^{\nu}c_i)f(x,y) \tag{7.104}$$

而在龙格-库塔公式的导出过程中，系数 $c_i(i=1,2,\cdots,\nu)$ 满足的第一个方程就是

$$c_1+c_2+\cdots+c_\nu=1$$

代入式(7.104)后得 $\varphi(x,y,0)=f(x,y)$，所以推知龙格-库塔方法是相容的方法。

若某单步法是 $p$ 阶的，$y(x)$ 为式(7.1)的解，则有

$$y(x+h)-y(x)=h\varphi(x,y,h)+O(h^{p+1})$$

$$\frac{y(x+h)-y(x)}{h}=\varphi(x,y,h)+\frac{O(h^{p+1})}{h} \tag{7.105}$$

由上式可见，若单步法式(7.91)与式(7.1)相容，则 $p$ 至少为 1。

## 6.3 单步法的绝对稳定性

关于单步法收敛的概念和收敛定理都是在计算过程无任何舍入误差的前提下建立起来的。整体截断误差 $\varepsilon_{n+1}=y(x_{n+1})-y_{n+1}$ 中的 $y_{n+1}$ 是以 $y_0$ 为起始值，由单步法式(7.91)经过精确计算得到的。但是实际计算时通常都会有舍入误差，特别是式(7.91)是一个递推算式，凡是递推算式都要考虑舍入误差的积累是否会得到扩大，也就是要考虑数值稳定性问题。一个单步法即使已满足相容条件，并且又是收敛的，然而在用它计算数值解时，如果舍入误差的积累越来越大，那么这样的数值方法也是没有实用价值的。

**定义 7.6** 对于固定的步长 $h$，如果用某一种单步法在计算 $y_n$ 时有大小为 $\varepsilon_n$ 的舍入误差，假定此后再没有新的舍入误差产生，若由 $\varepsilon_n$ 而引起的误差逐步衰减，则称此单步法是绝对稳定的。

设 $y_n$ 与 $\tilde{y}_n$ 为式(7.91)的准确值与带有舍入误差的值，则按式(7.91)计算时，前步误差对

本步计算结果所产生的误差可计算如下。记

$$y_{n+1}=y_n+h\varphi(x_n,y_n,h) \tag{7.106}$$

$$\tilde{y}_{n+1}=\tilde{y}_n+h\varphi(x_n,\tilde{y}_n,h) \tag{7.107}$$

$$\varepsilon_{n+1}=y_{n+1}-\tilde{y}_{n+1},\quad \varepsilon_n=y_n-\tilde{y}_n$$

则有

$$\begin{aligned}y_{n+1}-\tilde{y}_{n+1}&=(y_n-\tilde{y}_n)+h[\varphi(x_n,y_n,h)-\varphi(x_n,\tilde{y}_n,h)]\\&=(y_n-\tilde{y}_n)+h\frac{\partial\varphi(x_n,\tilde{y}_n+\theta\varepsilon_n,h)}{\partial y}(y_n-\tilde{y}_n)\\&=\left[1+h\frac{\partial\varphi(x_n,\tilde{y}_n+\theta\varepsilon_n,h)}{\partial y}\right](y_n-\tilde{y}_n)\end{aligned}$$

或

$$\varepsilon_{n+1}=\left[1+h\frac{\partial\varphi(x_n,\tilde{y}_n+\theta\varepsilon_n,h)}{\partial y}\right]\varepsilon_n,\qquad 0<\theta<1 \tag{7.108}$$

由于 $\varphi$ 与 $f$ 有关，当 $f$ 比较复杂时，误差方程(7.108)是一个非线性差分方程，其一般解自然不易求取。为了摆脱这种困难，经常将微分方程 $y'=f(x,y)$ 作线性化近似为更加简单的方程，以达到简化处理的目的，具体方法如下。

首先将 $f(x,y)$ 按台劳公式展开为

$$y'=f(x,y)=f(x_n,y_n)+\frac{\partial f}{\partial x}(x-x_n)+\frac{\partial f}{\partial y}(y-y_n)+\cdots$$

略去高于二次以上的项得

$$\begin{aligned}y'&=\frac{\partial f}{\partial y}y+\left[f(x_n,y_n)+\frac{\partial f}{\partial x}(x-x_n)-\frac{\partial f}{\partial y}y_n\right]\\&=\lambda y+\psi(x_n,y_n,x)\qquad\left(\lambda=\frac{\partial f}{\partial y}\right)\end{aligned}$$

或

$$y'(x)-\lambda y(x)=\psi(x_n,y_n,x) \tag{7.109}$$

对微分方程 $y'=f(x,y)$ 来说，式(7.109)是一个有效的近似。现将单步法式(7.106)和式(7.107)应用于方程(7.109)的求解，得到的是两个非齐次的差分方程。在一步求解的结果中，它们都含有对应于非齐次项 $\psi(x_n,y_n,x)$ 的相同特解，这个特解在误差方程(7.108)中被对消掉而不出现，所以它对以后各步的误差不产生影响。因此在下面的分析中，就将式(7.109)中的右端项略去，只就以下简化的方程

$$y'=\lambda y\qquad\left(\lambda=\frac{\partial f}{\partial y}\right) \tag{7.110}$$

来讨论单步法的数值稳定性问题。称式(7.110)为模型方程或试验方程，其中 $\lambda$ 一般可以取复数值，并设 $\mathrm{Re}(\lambda)<0$，以保证微分方程本身是稳定的。如果微分方程本身不稳定，则无论采用何种单步法，一般难以期望计算上的数值稳定性；如果微分方程本身是稳定的，则某些单步法是数值稳定的，另一些则是数值不稳定的。采用模型方程测试单步法的数值稳定性的直观理由是，如果单步法对这样简化了的方程是数值不稳定的，则对于更为复杂的方程也难以期望是数值稳定的。根据定义 7.6，采用模型方程测试单步法绝对稳定性的做法是，假定某步的数值解有舍入误差，而以后各步均没有舍入误差，这样来研究舍入误差的向下传播情况。如果用某一单步法解模型方程，则可得到以下差分方程

$$y_{n+1}=r(\lambda h)y_n \tag{7.111}$$

若设 $y_n$ 为单步法的准确值，$\tilde{y}_n$ 为具有舍入误差 $\varepsilon_n$ 的近似值，则有以下误差方程

$$\varepsilon_{n+1}=r(\lambda h)\varepsilon_n=r(\mu)\varepsilon_n \tag{7.112}$$

其中 $\mu=\lambda h$。如果要求单步法是绝对稳定的，就必须要求下式成立

$$|r(\mu)|<1 \tag{7.113}$$

**定义 7.7** 在复平面 $\mu$ 中，使 $|r(\mu)|<1$ 成立的区域，称为对应的单步法的绝对稳定区域。这个区域与实轴的交称为绝对稳定区间。如果某单步法的绝对稳定区域是有限的，则称该方法是条件稳定的；否则称为无条件稳定或恒稳定的。

绝对稳定区域反映了方法的稳定性对步长 $h$ 和 $\lambda$ 的限制。显然，绝对稳定区域越大，这个方法对 $h$ 的大小限制越小。在无条件稳定情况下，稳定性对步长 $h$ 的选取就并不构成限制。

对于欧拉公式，应用于模型方程(7.110)得

$$y_{n+1}=y_n+hf(x_n,y_n)=y_n+h\lambda y_n=(1+h\lambda)y_n=(1+\mu)y_n \tag{7.114}$$

误差方程为

$$\varepsilon_{n+1}=(1+\mu)\varepsilon_n \tag{7.115}$$

所以当

$$|1+\mu|<1 \tag{7.116}$$

时，欧拉方法是绝对稳定的。它的绝对稳定区域是一个以(−1,0)为中心的单位圆，如图 7.1 所示。它的绝对稳定区间是(−2,0)，所以当 $\lambda$ 为负实数时，取 $h<-2/\lambda$ 时，欧拉方法是绝对稳定的。

对于梯形公式，应用于模型方程得

$$y_{n+1}=y_n+\frac{h}{2}(\lambda y_{n+1}+\lambda y_n)$$

解得

$$y_{n+1}=\frac{1+\frac{1}{2}h\lambda}{1-\frac{1}{2}h\lambda}y_n=\frac{1+\frac{1}{2}\mu}{1-\frac{1}{2}\mu}y_n \tag{7.117}$$

误差方程为

$$\varepsilon_{n+1}=\frac{1+\frac{1}{2}\mu}{1-\frac{1}{2}\mu}\varepsilon_n \tag{7.118}$$

所以当

$$\left|\frac{1+\frac{1}{2}\mu}{1-\frac{1}{2}\mu}\right|<1 \tag{7.119}$$

时，梯形方法是绝对稳定的。式(7.118)当 $\mathrm{Re}(\lambda)<0$ 时恒稳定，故知它的绝对稳定区域是 $\mu$ 平面的整个左半平面(图 7.2)，它的稳定区域比欧拉法大得多，且没有有限的边界，属无条件稳定。

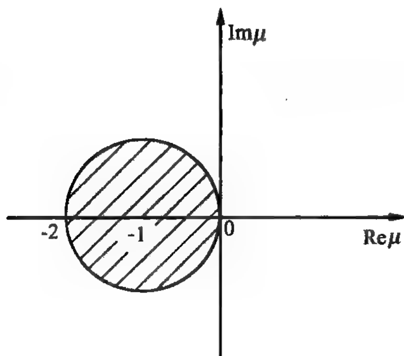



图 7.1

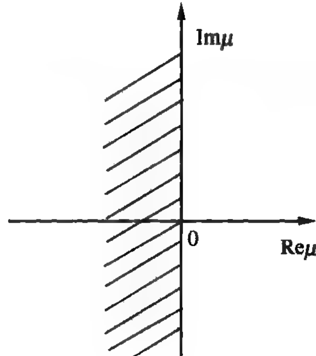



图 7.2

如果应用龙格-库塔法于模型方程,则有以下结果。

一阶龙格-库塔法

$$\begin{cases} y_{n+1}=y_n+k_1 \\ k_1=hf(x_n,y_n) \end{cases} \tag{7.120}$$

$$\begin{cases} |1+h\lambda|<1 & \text{(绝对稳定区域)} \\ (-2,0) & \text{(绝对稳定区间)} \end{cases} \tag{7.121}$$

二阶龙格-库塔法

$$\begin{cases} y_{n+1}=y_n+k_2 \\ k_1=hf(x_n,y_n) \\ k_2=hf(x_n+\frac{h}{2},y_n+\frac{k_1}{2}) \end{cases} \tag{7.122}$$

$$\begin{cases} \left|1+h\lambda+\frac{(h\lambda)^2}{2!}\right|<1 & \text{(绝对稳定区域)} \\ (-2,0) & \text{(绝对稳定区间)} \end{cases} \tag{7.123}$$

三阶龙格-库塔法

$$\begin{cases} y_{n+1}=y_n+\frac{1}{6}(k_1+4k_2+k_3) \\ k_1=hf(x_n,y_n) \\ k_2=hf(x_n+\frac{h}{2},y_n+\frac{k_1}{2}) \\ k_3=hf(x_n+h,y_n-k_1+2k_2) \end{cases} \tag{7.124}$$

$$\begin{cases} \left|1+h\lambda+\frac{(h\lambda)^2}{2!}+\frac{(h\lambda)^3}{3!}\right|<1 & \text{(绝对稳定区域)} \\ (-2.5,0) & \text{(绝对稳定区间)} \end{cases} \tag{7.125}$$

四阶龙格-库塔法

$$\begin{cases} y_{n+1}=y_n+\frac{1}{6}(k_1+2k_2+2k_3+k_4) \\ k_1=hf(x_n,y_n) \\ k_2=hf(x_n+\frac{h}{2},y_n+\frac{k_1}{2}) \\ k_3=hf(x_n+\frac{h}{2},y_n+\frac{k_2}{2}) \\ k_4=hf(x_n+h,y_n+k_3) \end{cases} \tag{7.126}$$

$$\begin{cases} \left|1+h\lambda+\frac{(h\lambda)^2}{2!}+\frac{(h\lambda)^3}{3!}+\frac{(h\lambda)^4}{4!}\right|<1 & \text{(绝对稳定区域)} \\ (-2.78,0) & \text{(绝对稳定区间)} \end{cases} \tag{7.127}$$

上述四种方法的绝对稳定区域见图 7.3,其中 $N$ 表示级数。从图中可见,龙格-库塔方法的绝对稳定区域随着 $N$ 的增大而扩大。可以证明 $m$ 阶龙格-库塔法应用于模型方程时的绝对稳定区域为

$$\left|1+h\lambda+\frac{(h\lambda)^2}{2!}+\cdots+\frac{(h\lambda)^m}{m!}\right|<1 \tag{7.128}$$

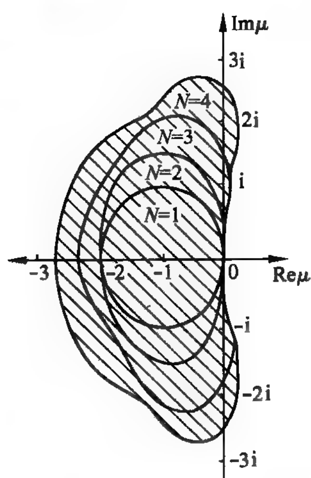



图 7.3

而且同阶的不同格式有相同的绝对稳定区域。

对于非线性微分方程 $y'=f(x,y)$，取 $\lambda=\partial f/\partial y$，此时 $\lambda$ 将是变化的，但只要在求解区间 $[a,b]$ 内，并且 $\mu=h\lambda$ 始终属于所用方法的绝对稳定区域内，则对此微分方程而言，该单步法是绝对稳定的。

## §7　差分方程简介

为讨论数值方法及与之有关理论的需要，下面对线性差分方程作简单的介绍。

具有常系数的线性差分方程的一般形式为

$$L(y(j))=a_0y_j+a_1y_{j+1}+\cdots+a_ny_{j+n}=c \tag{7.129}$$

不失一般性，可令 $j=0,1,2,\cdots$，其中 $a_0a_n\neq 0$，称式(7.129)为 $n$ 阶非齐次差分方程。而

$$L(y(j))=0 \tag{7.130}$$

称为齐次差分方程，下面首先给出齐次差分方程解的结构。为此作

$$P_n(x)=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_0 \tag{7.131}$$

$P_n(x)$ 称为差分方程(7.129)或(7.130)的特征方程。若 $x$ 为式(7.131)的某个根，令

$$y(j)=\alpha x^j \tag{7.132}$$

代入式(7.130)得

$$\begin{aligned}L(y(j))&=a_0(\alpha x^j)+a_1(\alpha x^{j+1})+\cdots+a_n(\alpha x^{j+n})\\&=\alpha x^j(a_0+a_1x+\cdots+a_nx^n)=0\end{aligned}$$

证得式(7.132)为齐次差分方程的解。

当 $x_1,x_2,\cdots,x_n$ 为特征方程(7.131)的几个相异根时，由差分方程的线性特性知下式

$$y(j)=\alpha_1x_1^j+\alpha_2x_2^j+\cdots+\alpha_nx_n^j \tag{7.133}$$

也满足齐次差分方程(7.130)，而对于给定的几个起始值 $y_0,y_1,\cdots,y_{n-1}$，代入式(7.133)后得以下线性方程组

$$\begin{cases}\alpha_1+\alpha_2+\cdots+\alpha_n=y_0\\ \alpha_1x_1+\alpha_2x_2+\cdots+\alpha_nx_n=y_1\\ \qquad\cdots\\ \alpha_1x_1^{n-1}+\alpha_2x_2^{n-1}+\cdots+\alpha_nx_n^{n-1}=y_{n-1}\end{cases} \tag{7.134}$$

由上述线性方程组可唯一地确定 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 的数值，它们可为实值或复值。

称 $\{y_k(j)\}_{k=1}^n=\{x_1^j,x_2^j,\cdots,x_n^j\}$ 为齐次差分方程的基本解集，基本解集是一组线性无关集，这组无关集的任何线性组合均为齐次差分方程的解；反之，差分方程的任何一组解均可表为这组基本解集的线性组合。这就是当特征方程有几个相异根时，齐次差分方程解的结构形式。

当 $x$ 为特征方程的 $r$ 重根时，容易验证以下各项

$$x^j,jx^j,j^2x^j,\cdots,j^{r-1}x^j \tag{7.135}$$

均是齐次差分方程的解，且它们是线性无关的解集。

设 $x_1,x_2,\cdots,x_m$ 分别为特征方程的 $n_1,n_2,\cdots,n_m$ 重根，这里 $\sum_{i=1}^{m} n_i=n$，则可得到齐次差分方程的一组线性无关解集

$$\begin{cases} x_1^j, jx_1^j, \cdots, j^{n_1-1}x_1^j \\ x_2^j, jx_2^j, \cdots, j^{n_2-1}x_2^j \\ \cdots \\ x_m^j, jx_m^j, \cdots, j^{n_m-1}x_m^j \end{cases} \tag{7.136}$$

这组线性无关解集的任何线性组合均能满足齐次差分方程(7.130)；反之，齐次差分方程(7.130)的任何一组解均可表为这组解集的线性组合。以上结论可写成

**定理 7.4**　若 $x_1,x_2,\cdots,x_m$ 为齐次差分方程(7.130)的特征方程

$$P_n(x)=\sum_{s=0}^{n} a_s x^s$$

的根，根的重数分别是 $n_1,n_2,\cdots,n_m$( $\sum_{i=1}^{m} n_i=n$)，则齐次差分方程的通解可表为

$$\begin{aligned} y(j)=&\alpha_{11}x_1^j+\alpha_{12}jx_1^j+\cdots+\alpha_{1n_1}j^{n_1-1}x_1^j+ \\ &\alpha_{21}x_2^j+\alpha_{22}jx_2^j+\cdots+\alpha_{2n_2}j^{n_2-1}x_2^j+\cdots+ \\ &\alpha_{m1}x_m^j+\alpha_{m2}jx_m^j+\cdots+\alpha_{mn_m}j^{n_m-1}x_m^j \\ =&\sum_{i=1}^{m}\sum_{l=1}^{n_i} a_{il}j^{l-1}x_i^j \end{aligned} \tag{7.137}$$

式(7.137)为齐次差分方程(7.130)的解的一般结构，而式(7.133)是其特例。若给定了 $n$ 个起始值 $y_0,y_1,\cdots,y_{n-1}$，分别把 $j=0,1,2,\cdots,n-1$ 代入式(7.137)可得到关于系数 $\alpha_{11},\alpha_{12},\cdots,\alpha_{1n_1};\alpha_{21},\alpha_{22},\cdots,\alpha_{2n_2};\cdots;\alpha_{m1},\alpha_{m2},\cdots,\alpha_{mn_m}$ 的一个 $n$ 阶线性方程组，由这个方程组，可唯一地确定这组系数。

由上可知，齐次差分方程的基本解集的表达式并不是唯一的。设齐次差分方程(7.130)对应于以下各组起始值

| $y_0$ | $y_1$ | $y_2$ | $\cdots$ | $y_{n-1}$ |
|---|---|---|---|---|
| 1 | 0 | 0 | $\cdots$ | 0 |
| 0 | 1 | 0 | $\cdots$ | 0 |
| $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ |
| 0 | 0 | 0 | $\cdots$ | 1 |

的解为

$$y(j)=y^{(k)}(j)\qquad (k=0,1,2,\cdots,n-1) \tag{7.138}$$

则式(7.138)也是齐次差分方程的一组基本解集，称为 $\delta$ 基本解集。而满足起始值为 $y_0=\eta_0$，$y_1=\eta_1,\cdots,y_{n-1}=\eta_{n-1}$ 的齐次差分方程(7.130)的通解为

$$y(j)=\eta_0 y^{(0)}(j)+\eta_1 y^{(1)}(j)+\cdots+\eta_{n-1}y^{(n-1)}(j) \tag{7.139}$$

对于非齐次差分方程(7.129)的解可由齐次差分方程的通解加上非齐次差分方程的一个特解得到。当 1 不是特征方程的根时，非齐次差分方程(7.129)有一特解为

$$\tilde{y}(j)=\frac{c}{\sum_{s=0}^{n} a_s} \tag{7.140}$$

**推论 1** 齐次差分方程的任何一组解 $y(j)(j=0,1,2,\cdots)$收敛于零的充分必要条件是，其特征方程 $P_n(x)=0$ 的根$|x_i|<1(i=1,2,\cdots,n)$。

**推论 2** 齐次差分方程的任何一组解 $y(j)(j=0,1,2,\cdots)$一致有界的充分必要条件是，其特征方程 $P_n(x)=0$ 满足根条件，即 $P_n(x)=0$ 的根$|x_i|\leqslant 1$，且落在单位圆边界上的根只能是单根。

# §8 线性多步法的相容性、收敛性与稳定性

## 8.1 线性多步法的相容性

线性多步法的相容性概念与单步法相同，即线性多步法的局部截断误差是否趋于零的问题。显然它只有当局部截断误差至少为 $h$ 的一阶无穷小量时才有可能。为使局部截断误差至少具有一阶量 $O(h)$，由式(7.54)可知，其充分必要条件是

$$c_0=0 \text{ 和 } c_1=0 \tag{7.141}$$

即

$$\sum_{j=0}^{k}\alpha_j=0 \text{ 和 } \sum_{j=0}^{k}j\alpha_j=\sum_{j=0}^{k}\beta_j \tag{7.142}$$

同时成立。记

$$\rho(\xi)=\sum_{j=0}^{k}\alpha_j\xi^j,\quad \sigma(\xi)=\sum_{j=0}^{k}\beta_j\xi^j \tag{7.143}$$

则式(7.142)等价于

$$\rho(1)=0 \text{ 和 } \rho'(1)=\sigma(1) \tag{7.144}$$

**定义 7.8** 线性多步法式(7.38)叫做与式(7.1)的微分方程相容，如果条件式(7.144)成立；并称条件式(7.144)为线性多步法式(7.38)的相容性条件。

## 8.2 线性多步法的收敛性

线性多步法的收敛性定义与单步法相同，它是否收敛与特征多项式 $\rho(\xi)$ 的根有关，为此引入以下定义。

**定义 7.9** 如果线性多步式(7.38)的特征多项式 $\rho(\xi)$ 的根都在单位圆内并且在单位圆上只有单根出现，则称线性多步法式(7.38)满足根条件。

**定理 7.5** 设线性多步法式(7.38)满足相容性条件和根条件，则当计算起始值的单步法收敛时，线性多步法式(7.38)也是收敛的；此外，若线性多步法式(7.38)是 $p$ 阶的，并且开始所用的单步法是不低于 $p$ 阶的，则该线性多步法的整体截断误差为

$$\varepsilon_{n+1}=y(x_{n+1})-y_{n+1}=O(h^p)\qquad(n+1\geqslant k) \tag{7.145}$$

证明略。

常用的线性多步法都满足相容条件和根条件。

## 8.3 线性多步法的绝对稳定性

研究线性多步法绝对稳定性的方法与单步法相同，将线性多步法式(7.38)用于求解模型方程 $y'=\lambda y$，得到计算公式

$$\sum_{j=0}^{k}(\alpha_j-\mu\beta_j)y_{n+j}=0\qquad(n=0,1,2,\cdots,M-k) \tag{7.146}$$

其中 $\mu=h\lambda$。设起始值为 $y_0,y_1,\cdots,y_{k-1}$；带有舍入误差的近似值为 $\tilde{y}_0,\tilde{y}_1,\cdots,\tilde{y}_{k-1}$，从这两组起始值出发，分别按式(7.146)精确计算可获得两组数值解 $y_m$ 与 $\tilde{y}_m(m\geqslant k)$，则误差 $\varepsilon_m=y_m-\tilde{y}_m$ 完全是由起始值的舍入误差引起的。于是有

$$\sum_{j=0}^{k}(\alpha_j-\mu\beta_j)\tilde{y}_{n+j}=0 \tag{7.147}$$

由(7.146)式减去(7.147)式得

$$\sum_{j=0}^{k}(\alpha_j-\mu\beta_j)\varepsilon_{n+j}=0 \tag{7.148}$$

(7.148)是一个常系数齐次差分方程，它的特征方程为

$$\sum_{j=0}^{k}(\alpha_j-\mu\beta_j)\xi^j=0 \tag{7.149}$$

或

$$\rho(\xi)-\mu\sigma(\xi)=0 \tag{7.150}$$

代数方程(7.149)或(7.150)也称为线性多步法式(7.38)的特征方程。设 $\xi_i$ 是特征方程(7.150)的 $r_i$ 重根($i=1,2,\cdots,s;r_1+r_2+\cdots+r_s=k$)，其中 $\xi_1,\xi_2,\cdots,\xi_s$ 互异，则差分方程(7.148)的通解为

$$\varepsilon_m=\sum_{i=1}^{s}\sum_{l=1}^{r_i}c_{il}m^{l-1}\xi_i^m \quad (m\geqslant k) \tag{7.151}$$

由初始误差 $\varepsilon_m=y_m-\tilde{y}_m(m=0,1,2,\cdots,k-1)$可定出 $k$ 个常数 $c_{il}$，并且每个 $c_{il}$ 都是 $\varepsilon_0,\varepsilon_1,\cdots,\varepsilon_{k-1}$ 的线性组合。由不等式

$$|\varepsilon_m|\leqslant\sum_{i=1}^{s}\sum_{l=1}^{r_i}|c_{il}|m^{l-1}|\xi_i|^m \tag{7.152}$$

可知，当特征方程(7.150)的所有根 $\xi_i$ 满足 $|\xi_i|<1$ 时，就有

$$\lim_{m\to\infty}\varepsilon_m=0$$

**定义 7.10**　对指定的 $\mu=h\lambda$，如果特征方程(7.150)的所有根 $\xi_i$ 按模小于 1，则称线性多步法(7.38)关于此 $\mu$ 是绝对稳定的。在复平面上所有使线性多步法为绝对稳定的 $\mu$ 的集合 $G$ 称为线性多步法式(7.38)的绝对稳定区域。

据上述定义，线性多步法式(7.38)的绝对稳定区域可表为

$$G=\left\{\mu\Big|_{|\xi|<1},\sum_{j=0}^{k}(\alpha_j-\mu\beta_j)\xi^j=0\right\} \tag{7.153}$$

与单步法的情形类似，如果求解的是一般非线性微分方程初值问题(7.1)，则应视 $\lambda=\partial f/\partial y$，这时 $\lambda$ 可能是变化的，但只要在求解区间$[a,b]$内，$\mu=h\lambda$ 始终属于绝对稳定区域 $G$，则对此微分方程而言，该线性多步法式(7.38)是绝对稳定的。

**例 7.8**　求显式米尔纳(Milne)法

$$y_{n+1}=y_{n-3}+\frac{h}{3}(8f_n-4f_{n-1}+8f_{n-2}) \tag{7.154}$$

的绝对稳定区域。

**解**　式(7.154)的特征方程为

$$\xi^1-\xi^{-3}-\frac{8}{3}\mu\xi^0+\frac{4}{3}\mu\xi^{-1}-\frac{8}{3}\mu\xi^{-2}=0$$

或

$$\xi^4-\frac{8}{3}\mu\xi^3+\frac{4}{3}\mu\xi^2-\frac{8}{3}\mu\xi-1=0 \tag{7.155}$$

方程(7.155)的所有根 $\xi_1,\xi_2,\xi_3,\xi_4$ 满足下式

$$|\xi_1\xi_2\xi_3\xi_4|=1$$

因而必存在某个根 $\xi_i$，使 $|\xi_i|\geqslant 1$，可见显式米尔纳方法不是一个绝对稳定的方法。

用直接求出特征方程的根去确定线性多步法的绝对稳定区域较为复杂，通常采用边界轨迹法。根据绝对稳定性的定义，就有 $\mu$ 复平面的绝对稳定区域 $G$ 与特征方程(7.150)的根位于 $\xi$ 复平面 $|\xi|<1$ 的单位圆域相对应。如果特征方程(7.150)的根位于单位圆周 $|\xi|=1$(即 $\xi=e^{i\theta}$)上，则满足方程

$$\rho(e^{i\theta})-\mu\sigma(e^{i\theta})=0 \qquad (0\leqslant\theta\leqslant 2\pi) \tag{7.156}$$

的 $\mu$ 就位于绝对稳定区域 $G$ 的边界上。因此，当 $\sigma(e^{i\theta})\neq 0(0\leqslant\theta\leqslant 2\pi)$ 时，边界曲线由

$$\mu(\theta)=\frac{\rho(e^{i\theta})}{\sigma(e^{i\theta})} \qquad (0\leqslant\theta\leqslant 2\pi) \tag{7.157}$$

即参数方程

$$\begin{cases} x(\theta)=\mathrm{Re}\mu(\theta) \\ y(\theta)=\mathrm{Im}\mu(\theta) \end{cases} \qquad (0\leqslant\theta\leqslant 2\pi)$$

给出。当 $\theta$ 从 0 变化到 $2\pi$ 时，$\mu(\theta)$ 在复平面上就描出边界曲线，该曲线把 $\mu$ 平面分为两个区域，如果在其中一个区域的内部存在一点 $\mu_1$，使方程 $\rho(\xi)-\mu_1\sigma(\xi)=0$ 有按模大于 1 的根，则该区域就不是绝对稳定区域；而另一个区域是绝对稳定区域，这就是边界轨迹法。

**例 7.9** 用边界轨迹法确定 $k=2$ 的显式阿当姆斯公式

$$y_{n+2}=y_{n+1}+\frac{h}{2}(3f_{n+1}-f_n) \tag{7.158}$$

的绝对稳定区域。

**解** 式(7.158)的特征方程为

$$\xi^2-\xi-\frac{\mu}{2}(3\xi-1)=0 \tag{7.159}$$

以 $\xi=e^{i\theta}=\cos\theta+i\sin\theta$ 代入上式解得

$$\mu(\theta)=\frac{2(e^{i2\theta}-e^{i\theta})}{3e^{i\theta}-1}=x(\theta)+iy(\theta)$$

其中

$$x(\theta)=-\frac{\cos 2\theta+4\cos\theta-3}{5-3\cos\theta}$$

$$y(\theta)=\frac{4\sin\theta-\sin 2\theta}{5-3\cos\theta}$$

由此取 $\theta_j(j=0,1,2,\cdots,N)$，算出 $(x(\theta_j),y(\theta_j))(j=0,1,2,\cdots,N)$ 并绘图，即可得到图 7.4 中 $k=2$ 的绝对稳定区域的边界曲线。在曲线内取一点 $(-\frac{2}{3},0)$，即 $\mu=-\frac{2}{3}$，代入式(7.159)得

$$\xi^2-\xi+\frac{1}{3}(3\xi-1)=\xi^2-\frac{1}{3}=0$$

显然有 $|\xi|=\sqrt{1/3}<1$，故该曲线围成的区域就是式(7.158)的绝对稳定区域。图 7.4 中 $k=3$，4 的显式阿当姆斯公式的绝对稳定区域也是用这种方法求出的。相应的绝对稳定区间列于表 7.7。

表 7.7

| 步数 $k$ | 阶 | 绝对稳定区间 |
|---|---|---|
| 1 | 1 | $(-2,0)$ |
| 2 | 2 | $(-1,0)$ |
| 3 | 3 | $(-\frac{6}{11},0)$ |
| 4 | 4 | $(-\frac{3}{10},0)$ |

同法可得隐式阿当姆斯公式的绝对稳定区域如图 7.5 所示，相应的绝对稳定区间列于表 7.8。

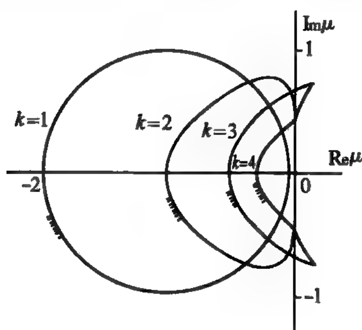



图 7.4

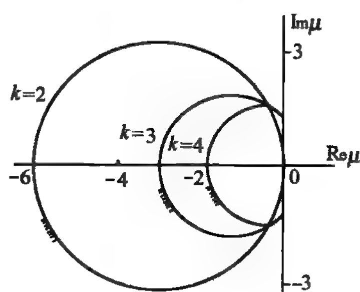



图 7.5

表 7.8

| 步数 $k$ | 阶 | 绝对稳定区间 |
|---|---|---|
| 1 | 2 | $(-\infty,0)$ |
| 2 | 3 | $(-6,0)$ |
| 3 | 4 | $(-3,0)$ |
| 4 | 5 | $(-\frac{90}{49},0)$ |

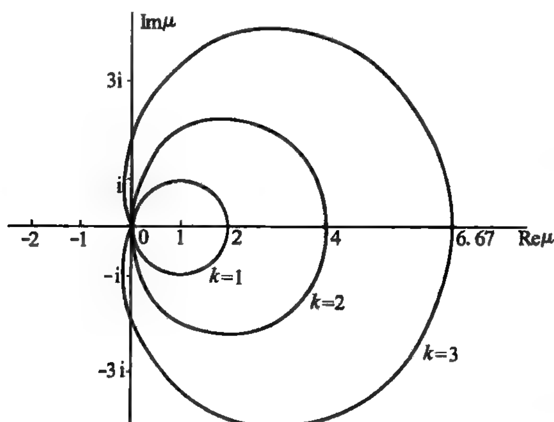



图 7.6

由上可见，对于相同阶的阿当姆斯公式，隐式公式的绝对稳定区域比显式公式大得多，且随着方法的阶数增高，其绝对稳定区域就变小。

利用边界轨迹法还可求出哈明方法(7.86)的绝对稳定区间为$(-8/3,0)$，$k$ 从 1 到 6 的吉尔方法的绝对稳定区域在图 7.6、图 7.7 绘出，图中区域的边界都是封闭曲线，绝对稳定区域是在相应闭曲线所围区域之外部。

**定义 7.11** 一个求解初值问题(7.1)的数值解法称为是 $A(\alpha)$-稳定的，如果它的绝对稳定区域包含了无限楔形区域(图 7.8)

$$W_\alpha=\{\mu \mid \pi-\alpha<\arg\mu<\pi+\alpha\} \tag{7.160}$$
$$0<\alpha\leqslant\frac{\pi}{2}$$

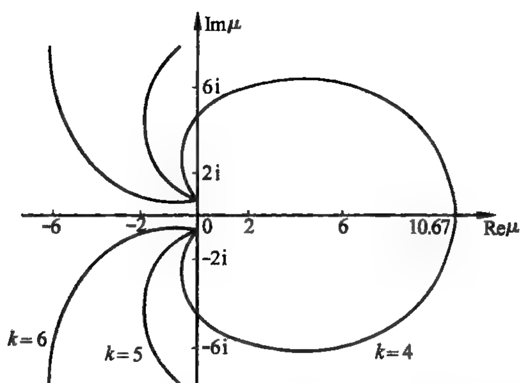


图 7.7

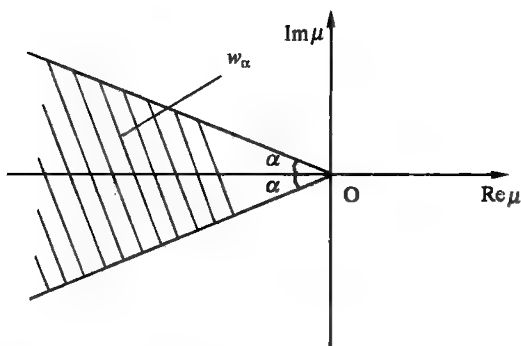


图 7.8

由图 7.6 和图 7.7 可见，$k=1$ 到 6 的吉尔方法都是 $A(\alpha)$-稳定的，其中每个 $\alpha$ 的最大值如表 7.9。

表 7.9

| $k$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\alpha_{\max}$ | 90° | 90° | 86°54′ | 73°14′ | 51°50′ | 18°47′ |

对于式(7.1)中的微分方程，只要 $\lambda=\partial f/\partial y$ 位于 $\mu$ 复平面的区域 $W_\alpha$ 内，那么单从数值稳定性考虑，使用 $A(\alpha)$-稳定的数值方法求解时，步长可不受限制，因为无论 $h>0$ 为何值，总有

$$\mu=h\lambda\in W_\alpha$$

**定义 7.12**　一个求解初值问题(7.1)的数值解法称为是 $A$-稳定的，如果它的绝对稳定区域包含了 $\mu$ 复平面的整个左半平面。

$A(\frac{\pi}{2})$-稳定就是 $A$-稳定的，例如梯形法和二步吉尔方法都是 $A$-稳定的方法。$A(\alpha)$-稳定的方法，其绝对稳定区域较 $A$-稳定的方法的绝对稳定区域小，因此若一个方法 $A$-稳定，则必然是 $A(\alpha)$-稳定。显然 $A$-稳定的数值方法同样对步长 $h$ 没有限制。

已经证明：显式线性多步法和显式龙格-库塔法都不可能是 $A$-稳定的；$A$-稳定的隐式线性多步法的阶不超过 2；在二阶隐式线性多步法中，误差常数最小的公式是梯形公式。

收敛性和稳定性是两个重要的概念，在数值分析的不同分枝，它们的含义可以不同。我们这里介绍的收敛性是反映数值方法本身的局部截断误差对计算结果的影响；稳定性是反映某一计算步骤中出现的舍入误差对计算结果的影响。只有既收敛又稳定的数值方法才能在实际计算中使用。

## §9　方法、阶和步长的选择

一个好的数值方法应该是用最少的工作量获得满足精度要求的计算结果。因此，除计算

公式的繁简特征外，误差是判别数值方法优劣的一个重要标志，这里只考虑截断误差是不够的，还必须考虑舍入误差。

求解初值问题(7.1)的任何一种数值方法，其计算量的大小主要取决于每步计算过程中需要对不同自变量计算函数 $f(x,y)$ 的次数。至于误差由以下两部分组成：$R_n=y(x_n)-y_n$ 是方法的整体截断误差；$\varepsilon_n=y_n-\tilde{y}_n$ 是舍入误差。其中 $y(x_n)$ 为微分方程的精确解，$y_n$ 为数值方法的精确解，$\tilde{y}_n$ 为数值方法的近似解。总误差为

$$E_n=y(x_n)-\tilde{y}_n=[y(x_n)-y_n]+[y_n-\tilde{y}_n]=R_n+\varepsilon_n \tag{7.161}$$

于是

$$|E_n|=|R_n+\varepsilon_n|\leqslant|R_n|+|\varepsilon_n| \tag{7.162}$$

$R_n$ 的大小与收敛性有关，鉴于整体截断误差比局部截断误差低一阶的关系，因此整体截断误差的大小一般采用对局部截断误差的精度要求来控制。$\varepsilon_n$ 的大小与计算的步数和数值稳定性有关，而数值稳定性又与微分方程的特性和采用的数值方法有关。在方法的选用上，应根据实际初值问题的特点和具体要求，结合不同类型数值方法的特性(见§12)进行选择。在方法择定的基础上，就可确定方法的阶。对于阶的选取还缺乏普遍有效的技术，如果视数值方法是对数值加工的工具，则一般处理的原则是：低精度问题采用粗加工工具，而高精度问题采用精加工工具，这种做法可使工作量最省。体现在数值计算中，那就是，对于式(7.1)的解光滑性较低或精度要求不高的情况，宜采用低阶方法；相反，应采用高阶方法。

在数值方法的阶已确定的情况下，选择合适的步长就是十分重要的工作。原则上，步长 $h$ 应由两个条件决定，一是要使求解过程绝对稳定；二是使方法的局部截断误差按模不超过给定的精度要求。对于第一个条件，只要根据所给初值问题和所用的数值方法的绝对稳定区域，就可确定出对 $h$ 的限制条件。对于第二个条件，由于局部截断误差估计的困难性，常采用事后估计误差法来自动选择满足精度要求的步长，这个方法对所有解初值问题的数值方法均适用，但通常仅限于龙格-库塔法，因为对于其他方法有更好的处理方法。

对于隐式方法，还必须考虑到迭代收敛性对步长 $h$ 的约束。如果 $h$ 较大，每步求解过程中的迭代次数就会增加，则在有限区间内，虽因 $h$ 增大使步数减少，但由此减少的工作量远不能抵消迭代过程增加的工作量。一般应选择合适的步长 $h$，使迭代次数不超过 2～3 次为宜。

从上面的分析可知，步长的选择与多种因素有关。从稳定性和局部截断误差着眼，步长 $h$ 愈小愈好。但 $h$ 太小，在一定范围内求解，步数就过多，这将导致舍入误差的严重积累和计算量的增加；而步长太大，又不能达到预期的精度要求或数值稳定性的要求。原则上，在满足稳定性及对局部截断误差要求的前提下，步长尽可能取大些。

## §10　常微分方程组和高阶微分方程的数值解法

一阶常微分方程组的初值问题如下：设有一阶常微分方程组

$$\begin{cases} y'_1=f_1(x,y_1,y_2,\cdots,y_m) \\ y'_2=f_2(x,y_1,y_2,\cdots,y_m) \\ \cdots \\ y'_m=f_m(x,y_1,y_2,\cdots,y_m) \end{cases} \quad (x\in[a,b]) \tag{7.163}$$

需要求取式(7.163)满足初始条件

$$y_i(x_0)=y_{i0} \qquad (i=1,2,\cdots,m) \tag{7.164}$$

的数值解，这里 $y_{i0}$ 为已知常数。这个问题称为微分方程组(7.163)满足初值条件的数值求解问题。

对于高阶微分方程，例如

$$y^{(m)}=f(x,y,y',y'',\cdots,y^{(m-1)}) \tag{7.165}$$

可令 $y_1=y, y_2=y'_1=y', y_3=y'',\cdots,y_m=y^{(m-1)}$，则式(7.165)可化为如下的一阶微分方程组

$$\begin{cases} y'_1=y_2 \\ y'_2=y_3 \\ \quad\cdots \\ y'_m=f(x,y_1,y_2,\cdots,y_m) \end{cases} \tag{7.166}$$

它是(7.163)的特殊情况。因此我们只须介绍一阶微分方程组的数值解法即可。

在方程组(7.163)中，$y_1(x),y_2(x),\cdots,y_m(x)$ 的个数 $m$ 称为微分方程组(7.163)的维数。记为

$$\boldsymbol{Y}=\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_m \end{bmatrix}, \quad \boldsymbol{Y}_0=\begin{bmatrix} y_{10} \\ y_{20} \\ \cdots \\ y_{m0} \end{bmatrix}, \quad \boldsymbol{F}=\begin{bmatrix} f_1 \\ f_2 \\ \cdots \\ f_m \end{bmatrix} \tag{7.167}$$

则初值问题式(7.163)、式(7.164)可表示成向量形式

$$\begin{cases} \boldsymbol{Y}'=\boldsymbol{F}(x,\boldsymbol{Y}) \\ \boldsymbol{Y}(x_0)=\boldsymbol{Y}_0 \end{cases} \quad x\in[a,b] \tag{7.168}$$

前面介绍过的一维情形的所有数值解法都可适用于 $n$ 维一阶联立微分方程组的情况。例如下列二维一阶联立微分方程组的初值问题

$$\begin{cases} y'=f(x,y,z) \\ z'=\varphi(x,y,z) \end{cases} \qquad \begin{cases} y(x_0)=y_0 \\ z(x_0)=z_0 \end{cases} \tag{7.169}$$

用古典龙格-库塔公式求解时，公式为

$$\begin{cases} y_{n+1}=y_n+\dfrac{1}{6}(k_{1n}+2k_{2n}+2k_{3n}+k_{4n}) \\ z_{n+1}=z_n+\dfrac{1}{6}(l_{1n}+2l_{2n}+2l_{3n}+l_{4n}) \end{cases} \quad (n=0,1,2,\cdots) \tag{7.170}$$

其中

$$\begin{cases} k_{1n}=hf(x_n,y_n,z_n) \\ k_{2n}=hf(x_n+\dfrac{h}{2},y_n+\dfrac{k_{1n}}{2},z_n+\dfrac{l_{1n}}{2}) \\ k_{3n}=hf(x_n+\dfrac{h}{2},y_n+\dfrac{k_{2n}}{2},z_n+\dfrac{l_{2n}}{2}) \\ k_{4n}=hf(x_n+h,y_n+k_{3n},z_n+l_{3n}) \end{cases} \qquad \begin{cases} l_{1n}=h\varphi(x_n,y_n,z_n) \\ l_{2n}=h\varphi(x_n+\dfrac{h}{2},y_n+\dfrac{k_{1n}}{2},z_n+\dfrac{l_{1n}}{2}) \\ l_{3n}=h\varphi(x_n+\dfrac{h}{2},y_n+\dfrac{k_{2n}}{2},z_n+\dfrac{l_{2n}}{2}) \\ l_{4n}=h\varphi(x_n+h,y_n+k_{3n},z_n+l_{3n}) \end{cases} \tag{7.171}$$

在二阶微分方程的情况下，古典公式的龙格-库塔法可以得到进一步的化简。设给定微分方程初值问题

$$y''=f(x,y,y'), \quad y(x_0)=y_0, \quad y'(x_0)=y_0' \tag{7.172}$$

令 $y'=z$，则可化为如下一阶联立微分方程组的初值问题：

$$\begin{cases} y'=z \\ z'=f(x,y,z) \end{cases} \qquad \begin{cases} y(x_0)=y_0 \\ z(x_0)=z_0 \end{cases} \tag{7.173}$$

此时相应的龙格-库塔（古典公式）的求解公式为

$$\begin{cases} y_{n+1}=y_n+\dfrac{1}{6}(k_{1n}+2k_{2n}+2k_{3n}+k_{4n}) \\ z_{n+1}=z_n+\dfrac{1}{6}(l_{1n}+2l_{2n}+2l_{3n}+l_{4n}) \end{cases} \quad (n=0,1,2,\cdots) \tag{7.174}$$

其中

$$\begin{cases} k_{1n}=hz_n \\ k_{2n}=h(z_n+\dfrac{l_{1n}}{2}) \\ k_{3n}=h(z_n+\dfrac{l_{2n}}{2}) \\ k_{4n}=h(z_n+l_{3n}) \end{cases} \qquad \begin{cases} l_{1n}=hf(x_n,y_n,z_n) \\ l_{2n}=hf(x_n+\dfrac{h}{2},y_n+\dfrac{k_{1n}}{2},z_n+\dfrac{l_{1n}}{2}) \\ l_{3n}=hf(x_n+\dfrac{h}{2},y_n+\dfrac{k_{2n}}{2},z_n+\dfrac{l_{2n}}{2}) \\ l_{4n}=hf(x_n+h,y_n+k_{3n},z_n+l_{3n}) \end{cases}$$

则式(7.174)可化为：

$$\begin{cases} y_{n+1}=y_n+h\left[y'_n+\dfrac{1}{6}(l_{1n}+l_{2n}+l_{3n})\right] \\ y'_{n+1}=y'_n+\dfrac{1}{6}(l_{1n}+2l_{2n}+2l_{3n}+l_{4n}) \end{cases} \quad (n=0,1,2,\cdots) \tag{7.175}$$

其中

$$\begin{cases} l_{1n}=hf(x_n,y_n,z_n) \\ l_{2n}=hf(x_n+\dfrac{h}{2},y_n+\dfrac{h}{2}z_n,z_n+\dfrac{l_{1n}}{2}) \\ l_{3n}=hf(x_n+\dfrac{h}{2},y_n+\dfrac{h}{2}z_n+\dfrac{h}{4}l_{1n},z_n+\dfrac{l_{2n}}{2}) \\ l_{4n}=hf(x_n+h,y_n+hz_n+\dfrac{h}{2}l_{2n},z_n+l_{3n}) \end{cases} \tag{7.176}$$

由一维情形建立的有关相容性和收敛性的概念和定理也适用于多维情形，此时应把函数的绝对值更换为函数向量的范数。例如，函数向量 $\boldsymbol{\Phi}(x,\boldsymbol{Y},h)$ 关于向量 $\boldsymbol{Y}$ 的李普希兹条件的形式为

$$\|\boldsymbol{\Phi}(x,\boldsymbol{Y}_1,h)-\boldsymbol{\Phi}(x,\boldsymbol{Y}_2,h)\|\leqslant L\|\boldsymbol{Y}_1-\boldsymbol{Y}_2\| \tag{7.177}$$

在微分方程组的情形，定义绝对稳定性的模型方程为 $m$ 维的线性微分方程

$$\boldsymbol{Y}'=\boldsymbol{\Lambda}\boldsymbol{Y} \tag{7.178}$$

式中，$\boldsymbol{\Lambda}$ 是对角矩阵

$$\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_m)$$

$\lambda_i(i=1,2,\cdots,m)$皆为复常数。

**定义 7.13**　设步长为 $h$ 的单步法(7.91)用于求解模型方程(7.178)的初值问题，并设初始向量$\boldsymbol{Y}_0$有误差$\boldsymbol{\varepsilon}_0=(\varepsilon_{10},\varepsilon_{20},\cdots,\varepsilon_{m0})'$，如果在计算后面的$\boldsymbol{Y}_n$时，由$\boldsymbol{\varepsilon}_0$所引起的误差$\boldsymbol{\varepsilon}_n=(\varepsilon_{1n},\varepsilon_{2n},\cdots,\varepsilon_{mn})'$满足

$$\lim_{n\to\infty}\boldsymbol{\varepsilon}_n=0$$

则称单步法(7.91)对于所用的步长 $h$ 和$\lambda_i(i=1,2,\cdots,m)$是绝对稳定的。

$m$ 维线性多步法的绝对稳定性定义与一维情形也完全相同。一维情形的各种数值解法的绝对稳定区域也是 $m$ 维情形下的绝对稳定区域。

例如,用欧拉法求模型方程(7.178)的初值问题,得计算公式

$$\boldsymbol{Y}_{n+1}=(\boldsymbol{I}+h\boldsymbol{\Lambda})\boldsymbol{Y}_n \tag{7.179}$$

其中 $\boldsymbol{I}$ 是 $m\times m$ 单位矩阵。设参加运算的实际数值为$\tilde{\boldsymbol{Y}}_n$,误差为 $\boldsymbol{\varepsilon}_n=\boldsymbol{Y}_n-\tilde{\boldsymbol{Y}}_n$,于是得

$$\tilde{\boldsymbol{Y}}_{n+1}=(\boldsymbol{I}+h\boldsymbol{\Lambda})\tilde{\boldsymbol{Y}}_n$$

上面两式相减得误差方程

$$\boldsymbol{\varepsilon}_{n+1}=(\boldsymbol{I}+h\boldsymbol{\Lambda})\boldsymbol{\varepsilon}_n$$

或

$$\varepsilon_{i(n+1)}=(1+h\lambda_i)\varepsilon_{in}\qquad(i=1,2,\cdots,m) \tag{7.180}$$

当

$$|1+h\lambda_i|<1\qquad(i=1,2,\cdots,m) \tag{7.181}$$

时,有

$$\lim_{n\to\infty}\boldsymbol{\varepsilon}_n=0$$

由此可知,$m$ 维欧拉法的绝对稳定区域仍是

$$|1+h\lambda|<1 \tag{7.182}$$

但是在使用时,$\lambda$ 要取遍 $\lambda_1,\lambda_2,\cdots,\lambda_m$。

设微分方程组(7.163)是一般常系数线性非齐次微分方程组

$$\boldsymbol{Y}'=\boldsymbol{A}\boldsymbol{Y}+\boldsymbol{\Psi}(x) \tag{7.183}$$

式中,$\boldsymbol{A}$ 是 $m\times m$ 常数矩阵,$\boldsymbol{\Psi}(x)=(\Psi_1(x),\Psi_2(x),\cdots,\Psi_m(x))'$。用欧拉法求解式(7.183)的初值问题,其计算公式是

$$\boldsymbol{Y}_{n+1}=\boldsymbol{Y}_n+h(\boldsymbol{A}\boldsymbol{Y}_n+\boldsymbol{\Psi}(x_n))=(\boldsymbol{I}+h\boldsymbol{A})\boldsymbol{Y}_n+h\boldsymbol{\Psi}(x_n)$$

令 $\boldsymbol{\varepsilon}_n=\boldsymbol{Y}_n-\tilde{\boldsymbol{Y}}_n$,因只考虑 $\boldsymbol{\varepsilon}_n$ 的传播,故有

$$\boldsymbol{\varepsilon}_{n+1}=(\boldsymbol{I}+h\boldsymbol{A})\boldsymbol{\varepsilon}_n \tag{7.184}$$

设 $A$ 有 $m$ 个线性无关的特征向量,则存在非奇异矩阵 $\boldsymbol{P}$,使

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_m)$$

或

$$\boldsymbol{A}=\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}$$

其中 $\lambda_i(i=1,2,\cdots,m)$是 $\boldsymbol{A}$ 的特征值。于是(7.184)式成为

$$\boldsymbol{\varepsilon}_{n+1}=(\boldsymbol{I}+h\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1})\boldsymbol{\varepsilon}_n=\boldsymbol{P}(\boldsymbol{I}+h\boldsymbol{\Lambda})\boldsymbol{P}^{-1}\boldsymbol{\varepsilon}_n$$

$$\boldsymbol{P}^{-1}\boldsymbol{\varepsilon}_{n+1}=(\boldsymbol{I}+h\boldsymbol{\Lambda})\boldsymbol{P}^{-1}\boldsymbol{\varepsilon}_n$$

或

$$\bar{\boldsymbol{\varepsilon}}_{n+1}=(\boldsymbol{I}+h\boldsymbol{\Lambda})\bar{\boldsymbol{\varepsilon}}_n$$

其中$\bar{\boldsymbol{\varepsilon}}_n=\boldsymbol{P}^{-1}\boldsymbol{\varepsilon}_n$,并且当$\lim\limits_{n\to\infty}\bar{\boldsymbol{\varepsilon}}_n=0$ 时也有$\lim\limits_{n\to\infty}\boldsymbol{\varepsilon}_n=0$。这时使用式(7.182)考察绝对稳定性时,不等式中的 $\lambda$ 也要取遍矩阵 $\boldsymbol{A}$ 的所有特征值。

设初值问题式(7.168)中的微分方程组是一般微分方程组,并设 $\boldsymbol{F}$ 关于 $\boldsymbol{Y}$ 的雅可比(Jacobi)矩阵

$$\frac{\partial\boldsymbol{F}}{\partial\boldsymbol{Y}}=\begin{bmatrix}\dfrac{\partial f_1}{\partial y_1} & \dfrac{\partial f_1}{\partial y_2} & \cdots & \dfrac{\partial f_1}{\partial y_m}\\ \dfrac{\partial f_2}{\partial y_1} & \dfrac{\partial f_2}{\partial y_2} & \cdots & \dfrac{\partial f_2}{\partial y_m}\\ \cdots & \cdots & \cdots & \cdots\\ \dfrac{\partial f_m}{\partial y_1} & \dfrac{\partial f_m}{\partial y_2} & \cdots & \dfrac{\partial f_m}{\partial y_m}\end{bmatrix} \tag{7.185}$$

在区间 $a \leqslant x \leqslant b$ 内有 $m$ 个线性无关的特征向量。那么在考察一个数值方法求解(7.168)的绝对稳定性时，该方法的绝对稳定区域中的 $\lambda$ 要取遍矩阵 $\partial F/\partial Y$ 的所有特征值。此时各个特征值 $\lambda$ 可能是变量，选取 $h$ 时，要使 $h\lambda$ 始终位于所用方法的绝对稳定区域内。

**例 7.10**　用欧拉法求解初值问题

$$\begin{cases} y' = xz - y + \sqrt{x} \\ z' = -2xy - 5z + \sin x \end{cases} \qquad \begin{cases} y(0) = y_0 \\ z(0) = z_0 \end{cases} \qquad (0 \leqslant x \leqslant 2)$$

时，分析绝对稳定性对步长有何限制?

**解**　因

$$\frac{\partial f_1}{\partial y} = -1, \quad \frac{\partial f_1}{\partial z} = x$$

$$\frac{\partial f_2}{\partial y} = -2x, \quad \frac{\partial f_2}{\partial z} = -5$$

得

$$\boldsymbol{A} = \begin{bmatrix} -1 & x \\ -2x & -5 \end{bmatrix}$$

由 $|\lambda \boldsymbol{I} - \boldsymbol{A}| = 0$ 解得 $\boldsymbol{A}$ 的特征值为

$$\lambda_{1,2} = -3 \pm \sqrt{4 - 2x^2}$$

当 $0 \leqslant x \leqslant \sqrt{2}$ 时，$\lambda_1, \lambda_2$ 都是实数，由欧拉法的绝对稳定区间条件 $-2 < h\lambda < 0$ 求解 $h$ 的取值范围如下：

由 $-2 < h\lambda_1 = h(-3 + \sqrt{4 - 2x^2}) < 0$ 解得

$$0 < h < \min_{0 \leqslant x \leqslant \sqrt{2}} \frac{2}{3 - \sqrt{4 - 2x^2}} = \frac{2}{3} = 0.\dot{6}$$

由 $-2 < h\lambda_2 < 0$ 解得

$$0 < h < \min_{0 \leqslant x \leqslant \sqrt{2}} \frac{2}{3 + \sqrt{4 - 2x^2}} = \frac{2}{5} = 0.5\dot{3}$$

由上可知，当 $0 \leqslant x \leqslant \sqrt{2}$ 时，对 $h$ 的限制为

$$0 < h < \frac{2}{5}$$

当 $\sqrt{2} < x \leqslant 2$ 时，$\lambda_{1,2} = -3 \pm \mathrm{i}\sqrt{2x^2 - 4}$。由 $-2 < h\lambda_{1,2} < 0$ 式求解 $h$ 的值域，取其实部得

$$0 < h < \min_{\sqrt{2} \leqslant x \leqslant 2} \frac{6}{2x^2 + 5} = \frac{6}{13} \approx 0.50$$

如果在整个求解区间 $0 \leqslant x \leqslant 2$ 内用不变的步长，则应使 $h$ 满足

$$0 < h < \frac{6}{13}$$

## §11　刚性方程组

在化学、自动控制、电力系统等领域中经常出现一类微分方程组，通常称为刚性方程组或 stiff 方程组。例如

$$\begin{cases} y'_1 = -0.1y_1 - 49.9y_2 \\ y'_2 = -50y_2 \\ y'_3 = 70y_2 - 120y_3 \end{cases} \tag{7.186}$$

及 $y_1(0)=2, y_2(0)=1, y_3(0)=2$，用矩阵形式可表为

$$\begin{bmatrix} y'_1 \\ y'_2 \\ y'_3 \end{bmatrix} = \begin{bmatrix} -0.1 & -49.9 & 0 \\ 0 & -50 & 0 \\ 0 & 70 & -120 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \tag{7.187}$$

其中

$$A = \begin{bmatrix} -0.1 & -49.9 & 0 \\ 0 & -50 & 0 \\ 0 & 70 & -120 \end{bmatrix}$$

由特征方程 $|\lambda I - A| = 0$ 求得 $\lambda_1 = -120, \lambda_2 = -50, \lambda_3 = -0.1$，则方程组(7.186)满足初值条件的理论解为

$$\begin{cases} y_1(x) = e^{-0.1x} + e^{-50x} \\ y_2(x) = e^{-50x} \\ y_3(x) = e^{-50x} + e^{-120x} \end{cases} \tag{7.188}$$

上述解一开始有一个短暂的变化较激烈的过渡过程，接着逐渐进入平稳状态。由(7.188)式可见，$y_1(x), y_2(x), y_3(x)$中相应于 $\lambda_1, \lambda_2$ 的项 $e^{-120x}, e^{-50x}$随 $x$ 的增加迅速地减小到可以忽略的程度。如果从数值方法的稳定性要求出发，$|h\lambda_i|\ (i=1,2,3)$都需要小于一个量，例如对欧拉法来说，这个量就是 2。如要求$|120h|<2$，因此 $h$ 的最大值只能为 1/60，这样确定的步长对于含有 $\lambda_1$ 的项来说，很快就没有实际价值了，但是在整个区间上，受绝对稳定性的限制，$h$ 又必须取得这样小，因而耗时较多且由于舍入误差的大量积累而使计算结果失真，因此用 $\lambda_1$ 来限制步长是不很自然的。但是，如果按 $\lambda_3$ 来选取步长，又不能使截断误差满足精度要求。基于上述原因，我们必须对这类特殊方程组考虑特殊的解法。

**定义 7.14** 若 $m$ 维常系数线性微分方程组

$$Y' = AY + \Psi(x) \tag{7.189}$$

的矩阵 $A$ 的特征值$\lambda_i\ (i=1,2,\cdots,m)$满足条件

① $\mathrm{Re}(\lambda_i) < 0 \quad (i=1,2,\cdots,m)$

② $\max\limits_i |\mathrm{Re}\lambda_i| \gg \min\limits_i |\mathrm{Re}\lambda_i|$

则称式(7.189)为刚性方程组，并称比值

$$r = \frac{\max\limits_i |\mathrm{Re}\lambda_i|}{\min\limits_i |\mathrm{Re}\lambda_i|} \tag{7.190}$$

为刚性比。对于一般 $m$ 维微分方程组(7.168)，如果它的雅可比矩阵式(7.185)的特征值 $\lambda_i\ (i=1,2,\cdots,m)$在区间 $a \leqslant x \leqslant b$ 上满足上述定义中的条件，那么，方程组(7.168)也称为刚性方程组。

用于刚性方程组的数值方法的绝对稳定性应当对 $h$ 不加限制，也就是说最好是 $A(\alpha)$—稳定或 $A$—稳定的方法。前面讲过的后退欧拉法、梯形法及各种吉尔方法就是这样一类用于求解刚性方程组的常用方法。如果使用 $A$—稳定的方法求解，则对任意的步长 $h>0, \mu = h\lambda_i\ (i=1,2,\cdots,m)$总位于方法的绝对稳定区域内，此时，只需从局部截断误差的控制考虑步长的选

择。如果使用 $A(\alpha)$—稳定的方法求解，那么，只要所有的 $\lambda_i$ 都位于所用方法的 $W_\alpha$ 区域内，则不论 $h>0$ 为何值，就总有 $\mu=h\lambda_i\in W_\alpha$。此时，同样只需从局部截断误差的控制考虑步长的选择。

到目前为止，$A$—稳定和 $A(\alpha)$—稳定的方法都是隐式方法，因此，对 $m$ 维的刚性方程组应用隐式线性 $k$ 步法求解时，每计算一步，都要用迭代法求解一个 $m$ 元非线性方程组，通常采用联立方程组的牛顿法求解。

在实际应用中，还有一类隐式非线性单步法亦可用于求解刚性方程组，它就是隐式龙格-库塔法。$N$ 级隐式龙格-库塔法的一般形式是

$$\begin{cases} y_{n+1} = y_n + \sum_{i=1}^{N} c_i k_i \\ k_i = hf(x_n + \alpha_i h, y_n + \sum_{j=1}^{N} b_{ij} k_j) \qquad (i=1,2,\cdots,N) \end{cases} \tag{7.191}$$

下面是三个常用的隐式龙格-库塔法。

① 一级二阶公式

$$\begin{cases} y_{n+1}=y_n+k_1 \\ k_1=hf(x_n+\frac{h}{2},y_n+\frac{h}{2}k_1) \end{cases} \tag{7.192}$$

② 二级二阶公式

$$\begin{cases} y_{n+1}=y_n+\frac{1}{2}(k_1+k_2) \\ k_1=hf(x_n,y_n) \\ k_2=hf(x_n+h,y_n+\frac{k_1+k_2}{2}) \end{cases} \tag{7.193}$$

③ 二级四阶公式

$$\begin{cases} y_{n+1}=y_n+\frac{k_1+k_2}{2} \\ k_1=hf\left(x_n+(\frac{1}{2}-\frac{\sqrt{3}}{6})h,y_n+\frac{k_1}{4}+(\frac{1}{4}-\frac{\sqrt{3}}{6})k_2\right) \\ k_2=hf\left(x_n+(\frac{1}{2}+\frac{\sqrt{3}}{6})h,y_n+(\frac{1}{4}+\frac{\sqrt{3}}{6})k_1+\frac{1}{4}k_2\right) \end{cases} \tag{7.194}$$

隐式龙格-库塔法是 $A$-稳定的，并且可以构造出 $N$ 级 $2N$ 阶隐式龙格-库塔方法。用 $N$ 级隐式龙格-库塔方法求解一个 $m$ 维微分方程组初值问题，每计算一步都要求解一个含 $mN$ 个未知量 $k_{ri}(r=1,2,\cdots,m;i=1,2,\cdots,N)$ 的非线性方程组，因此高级的隐式龙格-库塔方法在实际计算中不常用。

## §12　对各种方法的比较

考虑的解法类型有：① 台劳级数法；②龙格-库塔法；③显式线性多步法；④预测-校正法。下面就①、②、③、④四种类型解法，在各方面进行比较。

(1) 使用的简易性

主要考虑对计算机编程是否简单。

① 除低阶方法外，难以确定适当的公式。不需要特别的起始值计算程序(起始程序)。

② 非常简单。不要特别的起始程序。

③ 很直接，但要求特别的起始程序。

④ 除非每步仅有一次校正，实现时迭代增加了复杂性。它要求特别的起始程序。

(2) 计算量

常用每步中需要计算 $f$ 的次数来度量。如果 $f$ 是很复杂的，多数时间将花费在计算 $f$ 上。

① 每步要对 $f$ 及其导数作若干次计算，例如三阶方法需要六次计算。

② 计算 $f$ 次数较多，在四阶以内，计算的次数等于方法的阶数。

③ 计算 $f$ 一次，与阶无关。

④ 计算 $f$ 二次或三次，与阶无关。

(3) 局部截断误差

局部截断误差给出一类问题的方法精度的度量，但对一个特定的问题不能提供方法精度的正确比较。今考虑给定阶时局部截断误差的大小。

① 不很好。

② 可以很好。

③ 不如④。

④ 好。

(4) 局部截断误差估计

① 困难。

② 困难，但有某些方法给出局部截断误差的某些提示。

③ 困难。

④ 相当容易。

(5) 数值方法的绝对稳定性

不同类型以及同一类型不同阶的数值方法具有不同的绝对稳定区域，绝对稳定区域大者，其数值稳定性就越好。

(6) 步长调整

① 因是单步法，改变步长很容易。

② 因是单步法，改变步长很容易。

③ 很困难，需要特殊的公式。

④ 很困难，需要特殊的公式。

(7) 方法的选择

① 不推荐，除非问题有某些特点，例如，$f$ 及其导数由其他信息已经知道了。

② 可用于低精度的快速计算，亦可用于得到③和④的起始值。

③ 当 $f$ 很难计算时，可推荐这个方法。

④ 对多数可以估计局部截断误差的问题，推荐这个方法。

# 习 题 七

7.1 用欧拉方法计算下列初值问题的数值解

$$\begin{cases} y'=1+x\sin xy \\ y(0)=0,\ h=0.1 \qquad (0\leqslant x\leqslant 2) \end{cases}$$

7.2 用梯形法解 7.1。

7.3 用古典四阶龙格-库塔法求解下列初值问题

$$\begin{cases} y'=x+y \\ y(0)=1 \end{cases}$$

以步长 $h=0.1$ 计算 $y(0.1)$ 和 $y(0.2)$。

7.4 使用显式四阶阿当姆斯公式和隐式四阶阿当姆斯公式，分别求下列初值问题的数值解。

$$\begin{cases} y'=\dfrac{2}{x}y+x^2e^x, \qquad 1\leqslant x\leqslant 1.4 \\ y(1)=0,\ h=0.05 \end{cases}$$

7.5 用预测-校正法解初值问题

$$\begin{cases} y'=-5y, \qquad 0\leqslant x\leqslant 1 \\ y(0)=1, \quad h=0.1 \end{cases}$$

预测公式用显式四阶阿当姆斯公式计算；校正公式采用隐式四阶阿当姆斯公式计算；表头值用四阶龙格-库塔法计算。

7.6 将下列方程

$$y''-3y'+2y=0, y(0)=1, y'(0)=1$$

化为一阶联立微分方程组，并用古典四阶龙格-库塔法求解（求解区间[0,1]，取 $h=0.2$）。

7.7 用待定系数法确定如下求解公式的系数，使其阶数尽可能最高，并写出局部截断误差表达式。

(1) $y_{n+1}=\alpha_0 y_n+\alpha_1 y_{n-1}+\beta h f_{n+1}$

(2) $y_{n+1}=y_n+h(b_0 f_n+b_1 f_{n-1})$

(3) $y_{n+1}=\alpha y_{n-1}+h(bf_{n+1}+cf_n+df_{n-1})$

7.8 用二阶台劳展开法求初值问题

$$\begin{cases} y'=x^2+y^2 \\ y(1)=1 \end{cases}$$

的解在 $x=1.5$ 时的近似值（取 $h=0.25$，数值至少保留小数点后 5 位）。

7.9 求证以下求解初值问题的数值公式

$$y_{n+1}=y_n+\frac{h}{2}[3f(x_n,y_n)-f(x_{n-1},y_{n-1})]$$

是几阶公式？

7.10 将二阶微分方程初值问题 $y''=3y'-2y, y(0)=y'(0)=1$ 化为一阶微分方程组初值问题。

# 第八章　函数逼近

设 $C[a,b]$是在$[a,b]$上连续的全体函数构成的集合(或空间),对于 $f(x),g(x)\in C[a,b]$,用 $g(x)$来"逼近"$f(x)$时,必须明确"逼近"一词在数学上的确切含义,即明确如何度量 $f(x)$与 $g(x)$之间的"距离"。出于不同的理论与应用的目的,人们可以使用不同的"距离"尺度。回顾一下,在 $\mathbf{R}^n$ 空间中,曾对向量 $\boldsymbol{X}=(x_1,x_2,\cdots,x_n)$引入以下三种范数:

$$\|\boldsymbol{X}\|_1=\sum_{i=1}^{n}|x_i| \qquad \text{称为 1— 范数}$$

$$\|\boldsymbol{X}\|_2=(\sum_{i=1}^{n}x_i^2)^{\frac{1}{2}} \qquad \text{称为 2— 范数}$$

$$\|\boldsymbol{X}\|_\infty=\max_{1\leqslant i\leqslant n}|x_i| \qquad \text{称为 }\infty\text{— 范数}$$

利用它们,就可用 $\|\boldsymbol{X}-\boldsymbol{Y}\|$ 度量空间中两向量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 间的距离。

类似地,对 $f(x)\in C[a,b]$,也可以定义以下三种范数:

$$\|f\|_1=\int_a^b|f(x)|\mathrm{d}x \qquad \text{称为 1— 范数}$$

$$\|f\|_2=\left(\int_a^b f^2(x)\mathrm{d}x\right)^{\frac{1}{2}} \qquad \text{称为 2— 范数}$$

$$\|f\|_\infty=\max_{x\in[a,b]}|f(x)| \qquad \text{称为 }\infty\text{— 范数}$$

利用它们,就可用 $\|f-g\|$ 来度量两个函数 $f(x),g(x)$间的逼近程度。

本章所采用的逼近函数为 $m$ 次代数多项式 $P_m(x)$,采用的距离有以下两种。

(1) $$\|f(x)-P_m(x)\|_2=\left(\int_a^b[f(x)-P_m(x)]^2\mathrm{d}x\right)^{\frac{1}{2}}$$

这种度量方式也称平方度量。在离散化情况下,即 $f(x)$与 $P_m(x)$为定义在点集 $x_0,x_1,x_2,\cdots,x_n$ 上的函数时,上式右端演化为(即向量的欧几里得范数)

$$\left(\sum_{i=0}^{n}[f(x_i)-P_m(x_i)]^2\right)^{\frac{1}{2}}$$

在理论上或应用上,人们还经常用到加权平方度量,它给定一个权函数 $\rho(x)\geqslant 0$ 用来反映 $x$ 点处的差值$|f(x)-P_m(x)|$的重要程度,引入如下度量

$$\|f(x)-P_m(x)\|_2=\left(\int_a^b\rho(x)[f(x)-P_m(x)]^2\mathrm{d}x\right)^{\frac{1}{2}}$$

称为加权平方度量。

(2) $$\|f(x)-P_m(x)\|_\infty=\max_{x\in[a,b]}|f(x)-P_m(x)|$$

采用这种度量来刻画两函数间的逼近程度是很自然的,因为它关注的是两个函数在整个区间 $[a,b]$上的最大误差的大小,这种逼近称为一致逼近。

采用不同的范数来建立逼近准则,用于衡量和控制逼近误差,可获得不同的逼近函数。本章叙述以下两种逼近方法:最佳平方逼近(或最小平方逼近或最小二乘法)和最佳一致逼近(或一致逼近)。所采用的逼近函数以 $m$ 次多项式为主。

# §1　离散情况下的最小平方逼近

对于实验所获得的一组数据$(x_i,y_i)(i=0,1,2,\cdots,n)$，尽管它们之间存在函数关系，但其解析表达式并不知道。现在的问题是要用某种方法，找出能反映它们变化趋向的近似解析表达式$y=g(x)$，这种问题称为曲线拟合问题。用插值方法所找出的近似曲线，它通过所有已知点$(x_i,y_i)(i=0,1,2,\cdots,n)$，但是由观测或实验所获得的数据，不可避免地含有误差，这就使插值所找出的近似曲线保存所有观测误差，导致所得结果可能离开了实际情况，称为曲线拟合的过度现象。当$n$很大时，插值多项式的次数必然很高，这就大大增加了计算量；同时因舍入误差的增大，所得结果往往不可靠，因此要求拟合曲线通过所有点也未必好。而只要求尽可能地通过节点的函数值近旁，以部分地抵消原始数据中所包含的观测误差，从而使得结果能更好地反映客观实际，可以说这样做更具有实用价值，最小平方逼近就是其中的一个方法。

## 1.1　最小平方逼近问题概述

由前知，逼近问题可以区分为两类，其一是指在有限区间上的函数逼近；第二是函数的值仅在有限个点上已经限定的一类逼近，后者通常称为曲线拟合。一般情况下，采用的逼近函数亦有两类，一类是采用已知的基函数$\varphi_i(x)(i=0,1,2,\cdots,m)$的线性组合

$$g(x)=a_0\varphi_0(x)+a_1\varphi_1(x)+\cdots+a_m\varphi_m(x) \tag{8.1}$$

来逼近$y=f(x)$。等式(8.1)右端的线性组合称为线性数学模型。若以“最小平方逼近”为准则确定它，称这种类型的求解问题为线性最小平方逼近问题。其中基函数可能是$x$的非线性函数，但仅用参数$a_0,a_1,\cdots,a_m$作基函数的线性组合。例如具有固定指数的指数函数的如下组合

$$g(x)=a_0e^{\lambda_0x}+a_1e^{\lambda_1x}+\cdots+a_me^{\lambda_mx} \tag{8.2}$$

为线性数学模型；若$\lambda_i(i=0,1,2,\cdots,m)$也是待定参数，则它就成为非线性数学模型。最常见的线性数学模型如代数多项式

$$P_m(x)=a_0+a_1x+a_2x^2+\cdots+a_mx^m \tag{8.3}$$

和三角多项式

$$\begin{aligned}g(x)=a_0&+a_1\cos x+b_1\sin x+a_2\cos 2x+b_2\sin 2x+\cdots+\\&a_m\cos mx+b_m\sin mx\end{aligned} \tag{8.4}$$

如果函数$g(x)$的结构关于参数是非线性的，用最小平方逼近法求解这类问题称为非线性最小平方逼近问题。

## 1.2　方法描述

设有下述已知点组

$$(x_i,y_i)\qquad(i=0,1,2,\cdots,n)$$

今采用含有待定参数$a_0,a_1,a_2,\cdots,a_m(m\leqslant n)$的函数$g(x)$去拟合它，为使$g(x)$与数据点精确地吻合，这时选定参数问题成为一个“过确定”的问题（指方程个数大于未知数的个数），因而通常不可能使拟合函数与数据点精确地拟合，除非取$m=n$，这时拟合函数满足插值条件而成为插值函数。在确定参数$a_j(j=0,1,2,\cdots,m)$的许多不同准则中，最小平方逼近准则是最常用

的，方法如下。令

$$\varepsilon = g(x) - f(x)$$

为误差函数，建立

$$E = \| \varepsilon \|_2^2 = \sum_{i=0}^{n} \varepsilon_i^2 = \sum_{i=0}^{n} [g(x_i) - f(x_i)]^2 \tag{8.5}$$

称 $E$ 为 $g(x)$ 在 $n+1$ 个点 $x_i(i=0,1,2,\cdots,n)$ 上的误差平方和。因 $E$ 是 $a_0,a_1,\cdots,a_m$ 的连续函数，且 $E\geqslant 0$，所以一定存在一组数 $a_0,a_1,\cdots,a_m$ 使得 $E$ 取极小值。欲求 $E$ 的极小值，它应满足下列极值的必要条件

$$\frac{\partial E}{\partial a_j} = 0 \qquad (j = 0,1,2,\cdots,m) \tag{8.6}$$

当 $g(x)$ 为非线性数学模型时，联立方程组(8.6)是关于未知数 $a_0,a_1,\cdots,a_m$ 的一个非线性联立方程组，其求解比较困难。当 $g(x)$ 为线性数学模型时，按(8.6)式得

$$\frac{\partial}{\partial a_j} \sum_{i=0}^{n} \{[a_0\varphi_0(x_i) + a_1\varphi_1(x_i) + \cdots + a_m\varphi_m(x_i)] - f(x_i)\}^2$$

$$= 2\sum_{i=0}^{n} \{[a_0\varphi_0(x_i) + a_1\varphi_1(x_i) + \cdots + a_m\varphi_m(x_i)] - f(x_i)\} \cdot \varphi_j(x_i) = 0 \tag{8.7}$$

由此导得以下线性方程组

$$\begin{aligned} &\left(\sum_{i=0}^{n} \varphi_0(x_i)\varphi_j(x_i)\right)a_0 + \left(\sum_{i=0}^{n} \varphi_1(x_i)\varphi_j(x_i)\right)a_1 + \cdots + \left(\sum_{i=0}^{n} \varphi_m(x_i)\varphi_j(x_i)\right)a_m \\ &= \sum_{i=0}^{n} \varphi_j(x_i) f(x_i) \qquad (j = 0,1,2,\cdots,m) \end{aligned} \tag{8.8}$$

称(8.8)为正规方程组或法方程组。

### 1.3 关于法方程组解的唯一性及 $E$ 为极小值的证明

为了讨论法方程组解的唯一性，我们先引入以下有关的定义。

**定义 8.1** 一组函数 $\{\varphi_0(x),\varphi_1(x),\cdots,\varphi_m(x)\}$ 被称为在某个区间 $I$ 上是线性无关(或线性独立)的，是指如果对所有 $x\in I$，

$$\sum_{j=0}^{m} a_j\varphi_j(x) = 0 \tag{8.9}$$

当且仅当 $a_0=a_1=\cdots=a_m=0$；反之则说这组函数线性相关。

**定义 8.2** 设一组函数 $\{\varphi_0(x),\varphi_1(x),\cdots,\varphi_m(x)\}$ 中的每个函数都在区间 $I$ 上连续，如果对任意选取的不全为零的数 $a_0,a_1,\cdots,a_m$，函数

$$g(x) = \sum_{j=0}^{m} a_j\varphi_j(x) \tag{8.10}$$

在 $I$ 上具有的零点个数不多于 $m$ 个，则说该函数组是切比雪夫组(或者，等价地说它满足 Haar 条件)。

函数组 $\{\varphi_0(x),\varphi_1(x),\cdots,\varphi_m(x)\}$ 在 $I$ 上构成一个切比雪夫组的条件一定能保证该函数不含有线性相关的多余元素，因为如果能把某个 $\varphi_r(x)$ 表示成

$$\varphi_r(x) = \sum_{\substack{j=0 \\ j\neq r}}^{m} a_j\varphi_j(x) \qquad (\text{对于所有 } x \in I) \tag{8.11}$$

由此就可以推得式(8.10)中的 $g(x)$就会有多于 $m$ 个的零点(实际上,可推得有无限多个零点)。由此可知,一个函数组在 $I$ 上是切比雪夫组,可以肯定,它在 $I$ 上一定是线性独立的函数组;但是在 $I$ 上是线性独立的函数组,却未必一定是 $I$ 上的切比雪夫组。例如,函数组$\{1,x,x^3\}$在$[-1,+1]$上是线性独立的函数组,但它不是$[-1,+1]$上的切比雪夫组。因为函数 $g(x)=0\cdot 1+(-1)\cdot x+1\cdot x^3$在$[-1,+1]$上具有零点$-1,0,+1$,其零点数大于 2。

如果函数组$\{\varphi_0(x),\varphi_1(x),\cdots,\varphi_m(x)\}$在 $I$ 上的某一点集 $x_i(i=0,1,2,\cdots,n)$上线性相关,即存在非零系数 $r_j$ 使下式成立

$$\sum_{j=0}^{m} r_j\varphi_j(x_i) = 0 \qquad (i=0,1,2,\cdots,n) \tag{8.12}$$

那么 $r_j$ 的任何倍数都可以加到 $a_j$ 上而保持原误差的平方和不变,即有

$$\begin{aligned}\overline{E} &= \sum_{i=0}^{n}[(a_0+kr_0)\varphi_0(x_i)+\cdots+(a_m+kr_m)\varphi_m(x_i)-f(x_i)]^2 \\ &= \sum_{i=0}^{n}\{[a_0\varphi_0(x_i)+\cdots+a_m\varphi_m(x_i)-f(x_i)]+k[r_0\varphi_0(x_i)+\cdots+r_m\varphi_m(x_i)]\}^2 \\ &= \sum_{i=0}^{n}[a_0\varphi_0(x_i)+\cdots+a_m\varphi_m(x_i)-f(x_i)]^2 = E \end{aligned} \tag{8.13}$$

在这种情况下,按最小平方逼近法不一定能确定唯一的一组参数 $a_0,a_1,\cdots,a_m$。如果函数组$\{\varphi_0(x),\varphi_1(x),\cdots,\varphi_m(x)\}$在 $I$ 上构成一个切比雪夫组,即在 $I$ 上满足 Haar 条件,则这个条件既确保了上述函数组在单个点 $x\in I$ 上的线性独立性;同时又排除了上述函数组在 $I$ 的某一点集上可能出现的线性相关性。可以证明,在满足 Haar 条件下以下定理成立。

**定理 8.1** 如果函数组$\{\varphi_0(x),\varphi_1(x),\cdots,\varphi_m(x)\}$在包含点集 $x_i(i=0,1,2,\cdots,n)$的区间 $I$ 上满足 Haar 条件,则法方程组的系数矩阵非奇异,即最小平方逼近有唯一解 $a_0,a_1,\cdots,a_m$。

当法方程组有唯一解的情况下,留下的问题是,该解是否是 $E$ 的极小值点呢?可以指出,由法方程组获得的解确实能保证极小值。为确信这一点,考虑仅有两个函数 $\varphi_0(x)$和 $\varphi_1(x)$的情况,这时 $E$ 是$a_0$、$a_1$ 的函数,可表为 $E=E(a_0,a_1)$。令法方程组的解为 $a_0^*,a_1^*$,即它们满足

$$\frac{\partial E(a_0^*,a_1^*)}{\partial a_j} \equiv 0 \qquad (j=0,1)$$

考察差

$$\begin{aligned}&E(a_0^*+\delta_0,a_1^*+\delta_1)-E(a_0^*,a_1^*) \\ &=\sum_{i=0}^{n}\{[(a_0^*+\delta_0)\varphi_0(x_i)+(a_1^*+\delta_1)\varphi_1(x_i)]-f(x_i)\}^2- \\ &\quad\sum_{i=0}^{n}[a_0^*\varphi_0(x_i)+a_1^*\varphi_1(x_i)-f(x_i)]^2 \\ &=\sum_{i=0}^{n}[\delta_0\varphi_0(x_i)+\delta_1\varphi_1(x_i)]^2+2\delta_0\sum_{i=0}^{n}[a_0^*\varphi_0(x_i)+a_1^*\varphi_1(x_i)-f(x_i)]\cdot\varphi_0(x_i)+ \\ &\quad 2\delta_1\sum_{i=0}^{n}[a_0^*\varphi_0(x_i)+a_1^*\varphi_1(x_i)-f(x_i)]\cdot\varphi_1(x_i) \\ &=\sum_{i=0}^{n}[\delta_0\varphi_0(x_i)+\delta_1\varphi_1(x_i)]^2+2\delta_0\frac{\partial E(a_0^*,a_1^*)}{\partial a_0}+2\delta_1\frac{\partial E(a_0^*,a_1^*)}{\partial a_1}\end{aligned}$$

$$=\sum_{i=0}^{n}[\delta_0\varphi_0(x_i)+\delta_1\varphi_1(x_i)]^2\geqslant 0 \tag{8.14}$$

上式中的等号只有当

$$\sum_{i=0}^{n}[\delta_0\varphi_0(x_i)+\delta_1\varphi_1(x_i)]^2=0 \tag{8.15}$$

时才能达到。因 $\varphi_0(x)$ 与 $\varphi_1(x)$ 构成一个切比雪夫组,所以它们在点集 $x_i(i=0,1,2,\cdots,n)$ 上线性无关,除非 $\delta_0=\delta_1=0$,式(8.15)总是严格正的,这就证明了当 $a_0=a_0^*$ 和 $a_1=a_1^*$,$E(a_0,a_1)$ 确实取到了极小值 $E(a_0^*,a_1^*)$。对于一般的 $m<n$,可以类似地证明出

$$E(a_0^*+\delta_0,\ a_1^*+\delta_1,\cdots,a_m^*+\delta_m)-E(a_0^*,a_1^*,\cdots,a_m^*)\geqslant 0 \tag{8.16}$$

并当 $a_j=a_j^*\ (j=0,1,2,\cdots,m)$ 时,$E(a_0,a_1,\cdots,a_m)$ 取到极小值 $E(a_0^*,a_1^*,\cdots,a_m^*)$。

当 $\varphi_j(x)=x^j(j=0,1,2,\cdots,m)$ 时,最小平方逼近函数为下述最小平方逼近多项式

$$P_m(x)=a_0+a_1x+a_2x^2+\cdots+a_mx^m$$

其相应的法方程组为

$$\begin{cases}(n+1)a_0+(\sum_{i=0}^{n}x_i)a_1+(\sum_{i=0}^{n}x_i^2)a_2+\cdots+(\sum_{i=0}^{n}x_i^m)a_m=\sum_{i=0}^{n}f(x_i)\\ (\sum_{i=0}^{n}x_i)a_0+(\sum_{i=0}^{n}x_i^2)a_1+(\sum_{i=0}^{n}x_i^3)a_2+\cdots+(\sum_{i=0}^{n}x_i^{m+1})a_m=\sum_{i=0}^{n}x_if(x_i)\\ \cdots\\ (\sum_{i=0}^{n}x_i^m)a_0+(\sum_{i=0}^{n}x_i^{m+1})a_1+(\sum_{i=0}^{n}x_i^{m+2})a_2+\cdots+(\sum_{i=0}^{n}x_i^{2m})a_m=\sum_{i=0}^{n}x_i^mf(x_i)\end{cases} \tag{8.17}$$

上式亦可用矩阵形式表为

$$\boldsymbol{M}'\boldsymbol{M}\boldsymbol{A}=\boldsymbol{M}'\boldsymbol{Y} \tag{8.18}$$

其中

$$\boldsymbol{M}=\begin{bmatrix}1 & x_0 & x_0^2 & \cdots & x_0^m\\ 1 & x_1 & x_1^2 & \cdots & x_1^m\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 1 & x_n & x_n^2 & \cdots & x_n^m\end{bmatrix},\quad \boldsymbol{A}=\begin{bmatrix}a_0\\ a_1\\ \cdots\\ a_m\end{bmatrix},\quad \boldsymbol{Y}=\begin{bmatrix}f(x_0)\\ f(x_1)\\ \cdots\\ f(x_n)\end{bmatrix}=\begin{bmatrix}y_0\\ y_1\\ \cdots\\ y_n\end{bmatrix} \tag{8.19}$$

在满足 Haar 条件下,式(8.18)是一个对称正定的线性方程组,可采用平方根法求解之。

当 $m=0$ 时,即为零次最小平方逼近多项式 $P_0(x)=a_0$,其法方程式为

$$(n+1)a_0=\sum_{i=0}^{n}y_i$$

解得

$$a_0=\frac{y_0+y_1+\cdots+y_n}{n+1} \tag{8.20}$$

由上式可见,零次最小平方逼近多项式就是我们常用的平均值。

当 $m=1$ 时,即为一次最小平方逼近多项式 $P_1(x)=a_0+a_1x$,其法方程组为

$$\begin{cases}(n+1)a_0+(\sum x_i)a_1=\sum y_i\\ (\sum x_i)a_0+(\sum x_i^2)a_1=\sum x_iy_i\end{cases}$$

其中$\sum=\sum\limits_{i=0}^{n}$，解得

$$\begin{cases} a_0 = \dfrac{(\sum x_i^2)(\sum y_i)-(\sum x_i)(\sum x_i y_i)}{(n+1)(\sum x_i^2)-(\sum x_i)^2} \\ a_1 = \dfrac{(n+1)(\sum x_i y_i)-(\sum x_i)(\sum y_i)}{(n+1)(\sum x_i^2)-(\sum x_i)^2} \end{cases} \tag{8.21}$$

当$m=n$时，即为$n$次最小平方逼近多项式$P_n(x)$，式中的系数$a_j(j=0,1,2,\cdots,n)$可由

$$P_n(x_i) = y_i \qquad (i=0,1,2,\cdots,n) \tag{8.22}$$

单值确定，这时$E=0$。可见插值多项式是最小平方逼近多项式在$m=n$时的特殊情况。

**例 8.1** 有函数$y=f(x)$，具有函数值如表 8.1 所示，试用一个二次最小平方逼近多项式拟合它。

**解** 设二次最小平方逼近多项式为

$$P_2(x) = a_0 + a_1 x + a_2 x^2$$

表 8.1 中除$x_i, y_i$值外，还列入了法方程组中的有关数值。据表 8.1 的数据可建立法方程组如下

$$\begin{cases} 10a_0 + 4.5a_1 + 2.85a_2 = 31.761\,6 \\ 4.5a_0 + 2.85a_1 + 2.025a_2 = 14.089\,7 \\ 2.85a_0 + 2.025a_1 + 1.533\,3a_2 = 8.828\,8 \end{cases}$$

解得$a_0=3.195\,1$，　$a_1=0.442\,55$，　$a_2=-0.765\,31$。所以

$$P_2(x) = 3.195\,1 + 0.442\,55x - 0.765\,31x^2$$

**表 8.1**

| $i$ | $x_i$ | $y_i$ | $x_i^2$ | $x_i^3$ | $x_i^4$ | $x_i y_i$ | $x_i^2 y_i$ |
|---|---|---|---|---|---|---|---|
| 0 | 0.0 | 3.195 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0.1 | 3.229 9 | 0.01 | 0.001 | 0.000 1 | 0.322 99 | 0.032 299 |
| 2 | 0.2 | 3.253 2 | 0.04 | 0.008 | 0.001 6 | 0.650 64 | 0.130 128 |
| 3 | 0.3 | 3.261 1 | 0.09 | 0.027 | 0.008 1 | 0.978 33 | 0.293 499 |
| 4 | 0.4 | 3.251 6 | 0.16 | 0.064 | 0.025 6 | 1.300 64 | 0.520 255 |
| 5 | 0.5 | 3.228 2 | 0.25 | 0.125 | 0.062 5 | 1.614 10 | 0.807 050 |
| 6 | 0.6 | 3.180 7 | 0.36 | 0.216 | 0.129 6 | 1.908 42 | 1.145 052 |
| 7 | 0.7 | 3.126 6 | 0.49 | 0.343 | 0.240 1 | 2.188 62 | 1.532 034 |
| 8 | 0.8 | 3.059 4 | 0.64 | 0.512 | 0.409 6 | 2.446 52 | 1.958 016 |
| 9 | 0.9 | 2.975 9 | 0.81 | 0.729 | 0.656 1 | 2.678 31 | 2.410 179 |
| $\sum$ | 4.5 | 31.761 6 | 2.85 | 2.025 | 1.533 3 | 14.089 7 | 8.828 813 |

有时在解题时为了简化计算，可把已知数据用平移方法处理后再行计算，如下例。

**例 8.2** 数据表如表 8.2 所示，试用二次最小平方逼近多项式拟合它。

表 8.2

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $y$ | 15 | 14 | 14 | 14 | 14 | 15 | 16 |

**解** 设 $\bar{y}=y-14, \bar{x}=x-3$，则表 8.2 化为表 8.3。

表 8.3

| $\bar{x}$ | −3 | −2 | −1 | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| $\bar{y}$ | 1 | 0 | 0 | 0 | 0 | 1 | 2 |

这时

$$\sum \bar{x}_i = \sum \bar{x}_i^3 = 0$$

$$\sum \bar{x}_i^2 = 2(1^2+2^2+3^2) = 28$$

$$\sum \bar{x}_i^4 = 2(1^4+2^4+3^4) = 196$$

$$\sum \bar{y}_i = 4$$

$$\sum \bar{x}_i \bar{y}_i = 5$$

$$\sum \bar{x}_i^2 \bar{y}_i = 31$$

法方程组为

$$\begin{cases} 7a_0 + 28a_2 = 4 \\ 28a_1 = 5 \\ 28a_0 + 196a_2 = 31 \end{cases}$$

解得 $a_0=-\dfrac{1}{7}$， $a_2=a_1=\dfrac{5}{28}$。所以

$$\bar{y}=-\frac{1}{7}+\frac{5}{28}\bar{x}+\frac{5}{28}\bar{x}^2$$

或

$$y-14=-\frac{1}{7}+\frac{5}{28}(x-3)+\frac{5}{28}(x-3)^2$$

最小平方逼近有许多重要的应用，其中最简单的一个是关于过定方程组的求解。更多的是用在曲线拟合和函数逼近方面。

### 1.4 过定方程组的最小平方逼近解法

对于下列线性方程组

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i \qquad (i = 1,2,\cdots,m) \tag{8.23}$$

可用矩阵形式表示为

$$\boldsymbol{AX} = \boldsymbol{B} \tag{8.24}$$

其中

$$\boldsymbol{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \boldsymbol{X}=\begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}, \quad \boldsymbol{B}=\begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_m \end{bmatrix} \tag{8.25}$$

当 $m>n$ 时，称上述线性方程组为过定方程组或超定方程组。当秩 $r(\boldsymbol{A})<r(\boldsymbol{A}\vdots\boldsymbol{B})$ 时，方程组不相容，这时，又称为矛盾方程组，它无准确解，只能求近似解，经常采用最小平方逼近方法来求它的近似解。令

$$\varepsilon_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n - b_i$$

$$E = \sum_{i=1}^{n} \varepsilon_i^2$$

则 $E$ 取得极值的必要条件为

$$\frac{\partial E}{\partial x_j} = 0 \qquad (j = 1,2,\cdots,n)$$

由此获得以下法方程组

$$\boldsymbol{A}'\boldsymbol{A}\boldsymbol{X} = \boldsymbol{A}'\boldsymbol{B} \tag{8.26}$$

当 $\boldsymbol{A}$ 的秩为 $n$ 时，方程组(8.26)有唯一解；当 $\boldsymbol{A}$ 的秩 $<n$ 时，则有无穷多个解，它们均可作为过定方程组(8.23)的近似解。如超定方程组的秩 $r(\boldsymbol{A})=r(\boldsymbol{A}\cdots\boldsymbol{B})$，这时由法方程组(8.26)求得的解就是相容的方程组的解。如超定方程组不相容，则求得的解是超定方程组的近似解。实际问题中，一般 $m$、$n$ 都很大，且 $\boldsymbol{A}$ 中的 $a_{ij}$ 都有误差，按 $r(\boldsymbol{A})=r(\boldsymbol{A}\cdots\boldsymbol{B})$ 进行判别后求解未必切合实用。这时，采用法方程组(8.26)的求解方法既易于实现，又便于使用。

**例 8.3**　用最小平方逼近法求解下列超定方程组

$$\begin{cases} x_1 + x_2 + x_3 = 10 \\ x_1 + 3x_2 + 9x_3 = 5 \\ x_1 + 4x_2 + 16x_3 = 4 \\ x_1 + 5x_2 + 25x_3 = 2 \\ x_1 + 6x_2 + 36x_3 = 1 \\ x_1 + 7x_2 + 49x_3 = 1 \\ x_1 + 8x_2 + 64x_3 = 2 \end{cases}$$

**解**：记

$$\boldsymbol{A}=\begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \\ 1 & 6 & 36 \\ 1 & 7 & 49 \\ 1 & 8 & 64 \end{bmatrix}, \quad \boldsymbol{X}=\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \boldsymbol{B}=\begin{bmatrix} 10 \\ 5 \\ 4 \\ 2 \\ 1 \\ 1 \\ 2 \end{bmatrix}$$

按式(8.26)写出法方程组 $\boldsymbol{A}'\boldsymbol{A}\boldsymbol{X}=\boldsymbol{A}'\boldsymbol{B}$ 得

$$\begin{bmatrix} 7 & 34 & 200 \\ 34 & 200 & 1\,288 \\ 200 & 1\,288 & 8\,756 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 25 \\ 80 \\ 382 \end{bmatrix}$$

解得　$x_1=13.445\,1$，　$x_2=-3.585\,0$，　$x_3=0.263\,9$

## 1.5 可化为线性数学模型的最小平方拟合

### 1.5.1 变换关系列举

前面讨论的主要是线性数学模型的最小平方拟合问题。有些数学模型表面上不是线性数学模型,但通过变换可以化为线性数学模型。下面考虑常用的拟合函数:

$$\begin{cases} y = ax + b \\ y = ax^b \\ y = ab^x \\ y = a + \dfrac{b}{x} \\ y = \dfrac{1}{ax+b} \\ y = \dfrac{x}{ax+b} \\ y = a\ln x + b \end{cases} \tag{8.27}$$

对于上述拟合函数在表 8.4 的第 3 列中给出了转化为线性数学模型的有关变换关系,然后即可用线性数学模型的最小平方逼近法求解,再用逆变换得到原问题的拟合函数。

表 8.4

| 序号 | 拟合函数 | 变换后的形式 | $x_s$ | $y_s$ |
|---|---|---|---|---|
| 1 | $y=ax+b$ | | $\dfrac{x_0+x_n}{2}$ | $\dfrac{y_0+y_n}{2}$ |
| 2 | $y=ax^b$ | $Y=\alpha+bX$<br>$X=\lg x$<br>$Y=\lg y$<br>$\alpha=\lg a$ | $\sqrt{x_0x_n}$ | $\sqrt{y_0y_n}$ |
| 3 | $y=ab^x$ 或<br>$y=ae^{\beta x}$<br>$\beta=\ln b$ | $Y=\alpha+\beta x$<br>$Y=\lg y$<br>$\alpha=\lg a$<br>$\beta=\lg b$ | $\dfrac{x_0+x_n}{2}$ | $\sqrt{y_0\ y_n}$ |
| 4 | $y=a+\dfrac{b}{x}$ | $Y=ax+b$<br>$Y=xy$ | $\dfrac{2x_0x_n}{x_0+x_n}$ | $\dfrac{y_0+y_n}{2}$ |
| 5 | $y=\dfrac{1}{ax+b}$ | $Y=ax+b$<br>$Y=\dfrac{1}{y}$ | $\dfrac{x_0+x_n}{2}$ | $\dfrac{2y_0\ y_n}{y_0+y_n}$ |
| 6 | $y=\dfrac{x}{ax+b}$ | $Y=ax+b$<br>$Y=\dfrac{x}{y}$ | $\dfrac{2x_0x_n}{x_0+x_n}$ | $\dfrac{2y_0y_n}{y_0+y_n}$ |
| 7 | $y=a\lg x+b$ | $y=aX+b$<br>$X=\lg x$ | $\sqrt{x_0x_n}$ | $\dfrac{y_0+y_n}{2}$ |

### 1.5.2　拟合函数的选择

在最小平方拟合中，如何选择数学模型是很重要的。通常要根据物理意义或点组$(x_i,y_i)$ $(i=0,1,2\cdots,n)$的分布形状及特点去选择适当的拟合函数，并通过实际计算找出最好的拟合函数。

从数学上着眼，往往可以找到某种判据作为选择的依据。例如，对于一组实验数据，若采用多项式来拟合它，该多项式的次数可根据其某列差商（或差分）近似为常数的阶数取定。

对于表 8.2 中的拟合函数 $y=g(x,a,b)$，根据其凹向不变的特性，可以取定已知三点例如：$x_0<x_s<x_n$，若$(x_0,y_0)$、$(x_s,y_s)$、$(x_n,y_n)$位于曲线 $g(x,a,b)$上，则应有

$$y_0=g(x_0,a,b),\quad y_s=g(x_s,a,b),\quad y_n=g(x_n,a,b)$$

成立，消去 $a$、$b$ 后可获得 $y_s$ 值的计算公式。对 $x_s$ 值，可选定为 $x_s=\varphi(x_0,x_n)$的形式，相应地必对应有 $y_s=\psi(y_0,y_n)$的形式。为便于计算，要求 $x_s$、$y_s$ 具有比较简单的表达式。最后可取

$$\begin{cases}x_s=\varphi(x_0,x_n)\\ y_s=\psi(y_0,y_n)\end{cases}\tag{8.28}$$

作为$(x_s,y_s)$位于曲线 $g(x,a,b)$上的必要条件。今举下例说明之。

**例 8.4**　对下述幂函数

$$y=ax^b$$

建立必要条件(8.28)。

**解**　假定 $x_i>0,y_i>0(i=0,1,2,\cdots,n)$，如若不然，可作坐标平移变换后达到上述要求。在本例中，取定 $x_s$ 为 $x_0$ 与 $x_n$ 的几何平均值，即

$$x_s=\sqrt{x_0x_n}$$

则有

$$y_0=ax_0^b,\quad y_s=ax_0^{\frac{b}{2}}\cdot x_n^{\frac{b}{2}},\quad y_n=ax_n^b$$

消去 $a$、$b$ 后得

$$y_s=\sqrt{y_0\,y_n}$$

可见 $y_s$ 亦是对应于 $y_0$ 与 $y_n$ 的几何平均值。由此获得$(x_s,y_s)$位于曲线 $y=ax^b$ 上的必要条件为

$$\begin{cases}x_s=\sqrt{x_0x_n}\\ y_s=\sqrt{y_0y_n}\end{cases}$$

仿以上方法，可对表 8.4 中的每一个拟合函数建立其 $x_s$ 与 $y_s$ 的表达式，将它们列入 $x_s$、$y_s$ 列中以备查用。

为了从表 8.4 中选出吻合于点组$(x_i、y_i)(i=0,1,2,\cdots,n)$的拟合函数，可对每个拟合函数按其 $x_s$、$y_s$ 的表达式计算出它们的值。设 $x_s$ 介于 $x_i$ 与 $x_{i+1}$之间，使用线性插值公式

$$\hat{y}_s=y_i+\frac{y_{i+1}-y_i}{x_{i+1}-x_i}(x_s-x_i)$$

计算对应于 $x_s$ 的 $\hat{y}_s$ 值，建立差 $\Delta y_s=|y_s-\hat{y}_s|$，选出 $\Delta y_s$ 较小或次小者所对应的拟合函数作为数学模型，使用最小平方逼近法确定出公式中的参数 $a$、$b$，通过 $E$ 值的比较或根据拟合函数在点组$(x_i,y_i)(i=0,1,2,\cdots,n)$上的误差分布情况选定所需的拟合函数，这样求得的 $x$ 与 $y$ 间的函数关系通常称为经验公式。

从几何上说，表 8.4 中的拟合函数较适用于拟合具有单调性和凹凸性不变的实际函数。

表中所列必要条件仅是针对特定点$(x_i,y_i)(i=0,s,n)$建立的，即若$(x_0,y_0)$，$(x_s,y_s)$，$(x_n,y_n)$位于拟合函数上，就应有$x_s=\varphi(x_0,x_n)$、$y_s=\psi(y_0,y_n)$成立。由于实际数据都具有误差，因此按必要条件所得计算结果是近似的；另外，除上述三点外，没有考虑到其余点的分布性态，因而表 8.4 中所列必要条件仅是一种粗糙的判据。由于表 8.4 中的拟合函数均可通过变换转化为直线型关系的数学模型，因此可按变换后的数据点组，根据其与直线吻合的程度，选定拟合函数。这种方法能全面地反映出点组呈直线形分布的状况，且易于计算机处理，比较切合实用。表 8.4 中所列的只是最常用的一些拟合函数，另外，可能存在有不属于表 8.4 中的其他拟合函数，它们更吻合于给定数据点组的变化特性。

**例 8.5** 按表 8.5 选择拟合函数。

**表 8.5**

| $x$ | 273 | 283 | 288 | 293 | 313 | 333 | 353 | 373 |
|---|---|---|---|---|---|---|---|---|
| $y$ | 29.4 | 33.3 | 35.2 | 37.2 | 45.8 | 55.2 | 65.6 | 77.3 |

**解** 若仅限于式(8.27)中选择拟合函数，按表 8.4 中所列必要条件计算得数值表 8.6。由$\Delta y_s$列中的数值比较知，选择幂函数$y=ax^b$作为拟合函数较好。

**表 8.6**

| 序号 | $x_s$ | $y_s$ | $\hat{y}_s$ | $\Delta y_s$ | 拟合函数 |
|---|---|---|---|---|---|
| 1 | $\frac{x_0+x_n}{2}=\frac{273+373}{2}=323$ | $\frac{y_0+y_n}{2}=\frac{29.4+77.3}{2}=53.35$ | 50.5 | 2.85 | $y=ax+b$ |
| 2 | $\sqrt{x_0x_n}=\sqrt{273\times373}=319.1$ | $\sqrt{y_0y_n}=\sqrt{29.4\times77.3}=47.7$ | 48.7 | 1.0 | $y=ax^b$ |
| 3 | $\frac{x_0+x_n}{2}=323$ | $\sqrt{y_0y_n}=47.7$ | 50.5 | 2.8 | $y=ab^x$ |
| 4 | $\frac{2x_0x_n}{x_0+x_n}=\frac{2\times273\times373}{273+373}=315.3$ | $\frac{y_0+y_n}{2}=53.35$ | 46.9 | 6.45 | $y=a+\frac{b}{x}$ |
| 5 | $\frac{x_0+x_n}{2}=323$ | $\frac{2y_0y_n}{y_0+y_n}=\frac{2\times29.4\times77.3}{29.4+77.3}=42.6$ | 50.5 | 7.9 | $y=\frac{1}{ax+b}$ |
| 6 | $\frac{2x_0x_n}{x_0+x_n}=315.3$ | $\frac{2y_0y_n}{y_0+y_n}=42.6$ | 46.9 | 4.3 | $y=\frac{x}{ax+b}$ |
| 7 | $\sqrt{x_0x_n}=319.1$ | $\frac{y_0+y_n}{2}=53.35$ | 48.7 | 4.65 | $y=a\lg x+b$ |

## §2 离散情况下使用正交多项式的最小平方逼近

前面是用代数多项式$P_m(x)=\sum_{j=0}^{m}a_jx^j$在最小平方逼近准则下来拟合点组，虽然已指明正规方程组有唯一解，但在实际问题中，当$n$稍大时，已证明正规方程组是病态的，舍入误差会引起解的很大偏差，在计算机上用单字长求解时，其结果往往不太可靠。为避免上述弊端，可采用高精度运算；另一种方法就是采用一组正交多项式$\{P_k(x)\}_{k=0}^{m}$为基，用

$$Q_m(x)=\sum_{k=0}^{m}c_kP_k(x) \tag{8.29}$$

来拟合点组，其中 $P_k(x)$为 $k$ 次多项式，$c_k$ 为常数。这样可避免求解正规方程组，以保证求解的数值稳定性。

下面先介绍微积分和线性代数中未学过的新知识。

## 2.1　实轴上有限点系上的正交多项式

设有限点系为 $x_r(r=0,1,2,\cdots,n)$，今求多项式组 $P_i(x)(i=0,1,2,\cdots,m)$，使满足下列正交条件

$$\sum_{r=0}^{n}P_i(x_r)\cdot P_j(x_r)\begin{cases}\neq 0, & i=j\\ =0, & i\neq j\end{cases} \tag{8.30}$$

式中，$i,j\leqslant m<n$。为了确定这样的多项式组，我们用待定系数法计算 $P_i(x)$的递推关系式，设

$$\begin{cases}P_0(x)=1\\ P_1(x)=(x-\alpha_1)P_0(x)\\ P_{i+1}(x)=(x-\alpha_{i+1})P_i(x)-\beta_iP_{i-1}(x) \qquad (i=1,2,\cdots,m-1)\end{cases} \tag{8.31}$$

这里 $P_i(x)$是首项系数为 1 的 $i$ 次多项式。

将式(8.31)的第三式与 $P_i(x)$进行内积，由正交性得

$$\begin{aligned}0=(P_{i+1},P_i)&=((x-\alpha_{i+1})P_i,P_i)-\beta_i(P_{i-1},P_i)\\ &=(xP_i,P_i)-\alpha_{i+1}(P_i,P_i)\end{aligned}$$

所以
$$\alpha_{i+1}=\frac{(xP_i,P_i)}{(P_i,P_i)}=\frac{\sum\limits_{r=0}^{n}x_rP_i^2(x_r)}{\sum\limits_{r=0}^{n}P_i^2(x_r)} \tag{8.32}$$

再将式(8.31) 的第三式与 $P_{i-1}(x)$ 进行内积，由正交性得

$$\begin{aligned}0=(P_{i+1},P_{i-1})&=(xP_i,P_{i-1})-\alpha_{i+1}(P_i,P_{i-1})-\beta_i(P_{i-1},P_{i-1})\\ &=(xP_i,P_{i-1})-\beta_i(P_{i-1},P_{i-1})\end{aligned}$$

由内积定义知

$$(xP_i,P_{i-1})=(P_i,xP_{i-1})$$

且 $x\,P_{i-1}$是首项系数为 1 的 $i$ 次多项式，所以可表示为

$$xP_{i-1}=P_i+\sum_{j=0}^{i-1}c_j\,P_j$$

则有
$$(xP_{i-1},P_i)=(P_i,P_i)+\sum_{j=0}^{i-1}c_j(P_j,P_i)=(P_i,P_i)$$

从而得
$$\beta_i=\frac{(P_i,P_i)}{(P_{i-1},P_{i-1})}=\frac{\sum\limits_{r=0}^{n}P_i^2(x_r)}{\sum\limits_{r=0}^{n}P_{i-1}^2(x_r)} \tag{8.33}$$

这样就可按式(8.31)、式(8.32)、式(8.33) 构造有限点系上的正交多项式组。

对于一个 $m$ 次多项式 $Q_m(x)$ 可以用上述正交多项式 $P_0(x),P_1(x),\cdots,P_m(x)$ 的线性组合表为

$$Q_m(x) = c_0P_0(x) + c_1P_1(x) + \cdots + c_mP_m(x) \tag{8.34}$$

今使用 $Q_m(x)$来拟合点组$(x_i, y_i)(i=0,1,2,\cdots,n)$，要求找到系数 $c_0, c_1, \cdots, c_m$，使

$$I_m = \sum_{i=0}^{n}[Q_m(x_i) - y_i]^2 = I_m(c_0, c_1, \cdots, c_m) \tag{8.35}$$

取最小值。

## 2.2 $Q_m(x)$的求取方法

使式(8.35)取最小值的必要条件是 $c_j(j=0,1,2,\cdots,m)$满足

$$\frac{\partial I_m}{\partial c_j} = 0 \qquad (j = 0,1,2,\cdots,m) \tag{8.36}$$

即
$$2\sum_{i=0}^{n}[c_0P_0(x_i) + c_1P_1(x_i) + \cdots + c_mP_m(x_i) - y_i] \cdot P_j(x_i) = 0$$

利用正交性得

$$c_j\sum_{i=0}^{n}P_j^2(x_i) - \sum_{i=0}^{n}P_j(x_i)y_i = 0$$

解得
$$c_j = \frac{\sum_{i=0}^{n}P_j(x_i)y_i}{\sum_{i=0}^{n}P_j^2(x_i)} \qquad (j = 0,1,2,\cdots,m) \tag{8.37}$$

可以证明，这样求得的 $c_j(j=0,1,2,\cdots,m)$能使 $I_m$ 达到最小值。

综上可以给出求得 $Q_m(x)$拟合的计算步骤如下。

① 利用式(8.31)，式(8.32)，式(8.33)构造离散点系上的正交多项式组 $P_i(x)(i=0,1,2,\cdots,m)$。

② 按式(8.37)计算 $c_j(j=0,1,2,\cdots,m)$。

③ 写出 $Q_m(x)=c_0P_0(x)+c_1P_1(x)+\cdots+c_mP_m(x)$。

这里的 $m$ 可事先给定或在计算中根据误差确定。这种方法只需求内积运算而无须求解正规方程组，因而具有较好的数值稳定性。另外当 $Q_m(x)$精度不够而要建立新的 $Q_{m+1}(x)$时，只需在原来计算结果上增添一项 $c_{m+1}P_{m+1}(x)$即得，而不必从头算起，对计算机上使用的程序来说也只需把循环次数增 1 而其余不变。目前它是用多项式作曲线拟合的优良方法。

**例 8.6** 给定数据表表 8.7。

**表 8.7**

| $x_r$ | $-2$ | $-1$ | 0 | 1 | 2 |
|---|---|---|---|---|---|
| $y_r$ | $-1$ | $-1$ | 0 | 1 | 1 |

求最小平方逼近式 $Q_1(x), Q_2(x), Q_3(x)$。

**解** 按式(8.31)有

$$P_0(x) = 1$$
$$P_1(x) = (x - \alpha_1)P_0(x) = x - \alpha_1$$

$$\alpha_1=\frac{\sum_{r=0}^{4}x_rP_0^2(x_r)}{\sum_{r=0}^{4}P_0^2(x_r)}=\frac{\sum_{r=0}^{4}x_r}{5}=\frac{-2-1+0+1+2}{5}=0$$

得到

$$P_1(x)=x$$

$$P_2(x)=(x-\alpha_2)P_1(x)-\beta_1P_0(x)$$

$$\alpha_2=\frac{\sum_{r=0}^{4}x_r\,P_1^2(x_r)}{\sum_{r=0}^{4}P_1^2(x_r)}=\frac{\sum_{r=0}^{4}x_r^3}{\sum_{r=0}^{4}x_r^2}=0$$

$$\beta_1=\frac{\sum_{r=0}^{4}P_1^2(x_r)}{\sum_{r=0}^{4}P_0^2(x_r)}=\frac{\sum_{r=0}^{4}x_r^2}{5}=2$$

所以

$$P_2(x)=x^2-2$$

$$P_3(x)=(x-\alpha_3)P_2(x)-\beta_2P_1(x)$$

$$\alpha_3=\frac{\sum_{r=0}^{4}x_rP_2^2(x_r)}{\sum_{r=0}^{4}P_2^2(x_r)}=\frac{\sum_{r=0}^{4}x_r(x_r^2-2)^2}{\sum_{r=0}^{4}(x_r^2-2)^2}=0$$

$$\beta_2=\frac{\sum_{r=0}^{4}P_2^2(x_r)}{\sum_{r=0}^{4}P_1^2(x_r)}=\frac{\sum_{r=0}^{4}(x_r^2-2)^2}{\sum_{r=0}^{4}x_r^2}=\frac{7}{5}$$

得到

$$P_3(x)=(x-0)(x^2-2)-\frac{7}{5}x=x^3-\frac{17}{5}x$$

以下按式(8.37)计算得

$$c_0=\frac{\sum_{r=0}^{4}P_0(x_r)\cdot y_r}{\sum_{r=0}^{4}P_0^2(x_r)}=\frac{\sum_{r=0}^{4}y_r}{5}=0$$

$$c_1=\frac{\sum_{r=0}^{4}P_1(x_r)\cdot y_r}{\sum_{r=0}^{4}P_1^2(x_r)}=\frac{\sum_{r=0}^{4}x_ry_r}{\sum_{r=0}^{4}x_r^2}=0.6$$

$$c_2=\frac{\sum_{r=0}^{4}P_2(x_r)\cdot y_r}{\sum_{r=0}^{4}P_2^2(x_r)}=\frac{\sum_{r=0}^{4}(x_r^2-2)\cdot y_r}{\sum_{r=0}^{4}(x_r^2-2)^2}=0$$

$$c_3=\frac{\sum_{r=0}^{4}P_3(x_r)\cdot y_r}{\sum_{r=0}^{4}P_3^2(x_r)}=\frac{\sum_{r=0}^{4}(x_r^3-\frac{17}{5}x_r)\cdot y_r}{\sum_{r=0}^{4}(x_r^3-\frac{17}{5}x_r)^2}=-\frac{1}{6}$$

最后得

$$Q_1(x)=c_0P_0(x)+c_1P_1(x)=0.6x$$

$$Q_2(x)=c_0P_0(x)+c_1P_1(x)+c_2P_2(x)=Q_1(x)+c_2P_2(x)=0.6x$$

$$Q_3(x)=Q_2(x)+c_3P_3(x)=0.6x-\frac{1}{6}(x^3-\frac{17}{5}x)=\frac{1}{6}(7x-x^3)$$

## 2.3 多元函数的最小平方逼近

对于多变量函数离散情况下的最小平方逼近可以仿照单变量函数的逼近方法一样处理，为了避免求解高阶正规方程组，可以采用正交多项式组的最小平方逼近方法。以下介绍二元函数及三元函数的最小平方逼近方法及有关公式。

### 2.3.1 二元函数的最小平方逼近

设 $f(x,y)$ 是二元函数，如果在自变量点集 $(x_r,y_s)(r=0,1,2,\cdots,u;s=0,1,2,\cdots,v)$ 上的函数值是已知的，要求找出一个多项式

$$Q_{m,n}(x,y)=\sum_{i=0}^{m}\sum_{j=0}^{n}a_{ij}P_i(x)q_j(y) \tag{8.38}$$

使得多元函数 $I_{m,n}$ 取到最小值，即

$$I_{m,n}=\sum_{r=0}^{u}\sum_{s=0}^{v}[f(x_r,y_s)-Q_{m,n}(x_r,y_s)]^2=\min \tag{8.39}$$

式中，$a_{ij}$ 为系数，$P_i(x),q_j(y)$ 为满足下列条件的正交多项式

$$\sum_{r=0}^{u}P_i(x_r)P_k(x_r)\begin{cases}\neq 0, & i=k\\ =0, & i\neq k\end{cases}$$

$$\sum_{s=0}^{v}q_j(y_s)q_k(y_s)\begin{cases}\neq 0, & j=k\\ =0, & j\neq k\end{cases} \tag{8.40}$$

使式(8.39)成立的 $a_{hk}$ 应满足

$$\frac{\partial I_{m,n}}{\partial a_{hk}}=0 \qquad (h=0,1,2,\cdots,m;\quad k=0,1,2,\cdots,n)$$

即

$$\begin{aligned}
&\frac{\partial}{\partial a_{hk}}\sum_{r=0}^{u}\sum_{s=0}^{v}\Big[f(x_r,y_s)-\sum_{i=0}^{m}\sum_{j=0}^{n}a_{ij}P_i(x_r)q_j(y_s)\Big]^2\\
&=2\sum_{r=0}^{u}\sum_{s=0}^{v}\Big[f(x_r,y_s)-\sum_{i=0}^{m}\sum_{j=0}^{n}a_{ij}P_i(x_r)q_j(y_s)\Big]\cdot P_h(x_r)q_k(y_s)\\
&=2\Big\{\sum_{r=0}^{u}\sum_{s=0}^{v}f(x_r,y_s)P_h(x_r)q_k(y_s)-\sum_{i=0}^{m}\sum_{j=0}^{n}a_{ij}\Big[\sum_{r=0}^{u}P_i(x_r)P_h(x_r)\Big]\Big[\sum_{s=0}^{v}q_j(y_s)q_k(y_s)\Big]\Big\}\\
&=2\Big\{\sum_{r=0}^{u}\sum_{s=0}^{v}f(x_r,y_s)P_h(x_r)q_k(y_s)-a_{hk}\Big[\sum_{r=0}^{u}P_h^2(x_r)\Big]\Big[\sum_{s=0}^{v}q_k^2(y_s)\Big]\Big\}=0
\end{aligned} \tag{8.41}$$

因此得

$$a_{hk}=\frac{\sum_{r=0}^{u}\sum_{s=0}^{v}f(x_r,y_s)P_h(x_r)q_k(y_s)}{\Big[\sum_{r=0}^{u}P_h^2(x_r)\Big]\Big[\sum_{s=0}^{v}q_k^2(y_s)\Big]} \qquad (h=0,1,2,\cdots,m;\quad k=0,1,2,\cdots,n) \tag{8.42}$$

#### 2.3.2 三元函数的最小平方逼近

设 $f(x,y,z)$是三元函数，如果在自变量点集$(x_r,y_s,z_t)$($r=0,1,2,\cdots,u$;$s=0,1,2,\cdots,v$;$t=0,1,2,\cdots,w$)上的函数值是已知的，要求找出一个多项式

$$Q_{m,n,l}(x,y,z)=\sum_{i=0}^{m}\sum_{j=0}^{n}\sum_{k=0}^{l}a_{ijk}P_i(x)q_j(y)r_k(z) \tag{8.43}$$

使得多元函数 $I_{m,n,l}$取到最小值，即

$$I_{m,n,l}=\sum_{r=0}^{u}\sum_{s=0}^{v}\sum_{k=0}^{w}[f(x_r,y_s,z_t)-Q_{m,n,l}(x_r,y_s,z_t)]^2=\min \tag{8.44}$$

其中 $P_i(x)$、$q_j(y)$、$r_k(z)$分别是点集$\{x_r\}$、$\{y_s\}$、$\{z_t\}$上的 $i$ 次、$j$ 次、$k$ 次正交多项式。用同样方法可得到满足式(8.44)的系数值 $a_{ijk}$ 为

$$a_{ijk}=\frac{\sum\limits_{r=0}^{u}\sum\limits_{s=0}^{v}\sum\limits_{t=0}^{w}f(x_r,y_s,z_t)P_i(x_r)q_j(y_s)r_k(z_t)}{\left[\sum\limits_{r=0}^{u}P_i^2(x_r)\right]\left[\sum\limits_{s=0}^{v}q_j^2(y_s)\right]\left[\sum\limits_{t=0}^{w}r_k^2(z_t)\right]}$$

$$(i=0,1,2,\cdots,m;\quad j=0,1,2,\cdots,n;\quad k=0,1,2,\cdots,l) \tag{8.45}$$

更多自变量的多元函数的最小平方逼近多项式可仿上推导。

## §3 连续情况下的最小平方逼近

对于$[a,b]$上给定的函数 $f(x)$用 $m$ 次多项式

$$P_m(x)=\sum_{i=0}^{m}a_ix^i \tag{8.46}$$

近似它，要求多元函数 $I_m(a_0,a_1,a_2,\cdots,a_m)$取到最小值，即

$$I_m=\int_a^b[P_m(x)-f(x)]^2\,\mathrm{d}x=\min \tag{8.47}$$

其正规方程组为

$$\frac{\partial I_m}{\partial a_j}=2\int_a^b\left[\sum_{i=0}^{m}a_ix^i-f(x)\right]\cdot x^j\,\mathrm{d}x=0\qquad(j=0,1,2,\cdots,m) \tag{8.48}$$

为了避免求解正规方程组，可以采用在$[a,b]$上正交的多项式组 $P_i(x)$($i=0,1,2,\cdots m$)的线性组合作为近似多项式

$$Q_m(x)=\sum_{i=0}^{m}c_iP_i(x) \tag{8.49}$$

其中 $P_i(x)$ ($i=0,1,2,\cdots,m$)满足下列正交条件

$$\int_a^bP_k(x)P_l(x)\mathrm{d}x\begin{cases}\neq 0, & k=l\\ =0, & k\neq l\end{cases} \tag{8.50}$$

则 $Q_m(x)$作为 $f(x)$在区间$[a,b]$上的最小平方逼近多项式时的有关公式，只需将式(8.30)～式(8.37)中的求和符号改为$[a,b]$上的求积符号，即得

$$I_m=\int_a^b[Q_m(x)-f(x)]^2\mathrm{d}x=\min \tag{8.51}$$

$$\begin{cases}P_0(x)=1\\ P_1(x)=(x-\alpha_1)P_0(x)\\ P_{i+1}(x)=(x-\alpha_{i+1})P_i(x)-\beta_iP_{i-1}(x)\qquad(i=1,2,\cdots,m-1)\end{cases} \tag{8.52}$$

其中

$$\alpha_{i+1}=\frac{\int_a^b xP_i^2(x)\mathrm{d}x}{\int_a^b P_i^2(x)\mathrm{d}x} \tag{8.53}$$

$$\beta_i=\frac{\int_a^b P_i^2(x)\mathrm{d}x}{\int_a^b P_{i-1}^2(x)\mathrm{d}x} \tag{8.54}$$

$$c_i=\frac{\int_a^b P_i(x)\ f(x)\mathrm{d}x}{\int_a^b P_i^2(x)\mathrm{d}x} \tag{8.55}$$

我们亦可利用一些常用的经典的正交多项式组来建立 $Q_m(x)$。如定义在$[-1,+1]$上的勒让德正交多项式

$$\mathrm{P}_n(x)=\frac{1}{2^n n!}\frac{\mathrm{d}^n}{\mathrm{d}x^n}(x^2-1)^n \tag{8.56}$$

它的各次多项式按以下递推公式导出

$$\mathrm{P}_{n+1}(x)=\frac{2n+1}{n+1}x\mathrm{P}_n(x)-\frac{n}{n+1}\mathrm{P}_{n-1}(x) \tag{8.57}$$

其前几个多项式为

$$\begin{cases}\mathrm{P}_0(x)=1\\ \mathrm{P}_1(x)=x\\ \mathrm{P}_2(x)=\frac{1}{2}(3x^2-1)\\ \mathrm{P}_3(x)=\frac{1}{2}(5x^3-3x)\\ \mathrm{P}_4(x)=\frac{1}{8}(35x^4-30x^2+3)\\ \mathrm{P}_5(x)=\frac{1}{8}(63x^5-70x^3+15x)\end{cases} \tag{8.58}$$

这些多项式满足以下正交条件

$$\int_{-1}^{+1}\mathrm{P}_n(x)\mathrm{P}_m(x)\mathrm{d}x=\begin{cases}\frac{2}{2n+1}, & m=n=0,1,2,\cdots\\ 0, & m\neq n\end{cases} \tag{8.59}$$

由于 $\mathrm{P}_n(-x)=(-1)^n\mathrm{P}_n(x)$，所以偶次勒让德多项式是偶函数；奇次勒让德多项式是奇函数。

对于在区间$[a,b]$上的函数 $f(z)$，可引入线性变换

$$z=\frac{b-a}{2}x+\frac{b+a}{2}\qquad(-1\leqslant x\leqslant+1) \tag{8.60}$$

用

$$x=\left(z-\frac{b+a}{2}\right)\Big/\left(\frac{b-a}{2}\right)$$

代入勒让德多项式组后得

$$\widetilde{\mathrm{P}}_i(z)=\mathrm{P}_i\left(\frac{z-\frac{b+a}{2}}{\frac{b-a}{2}}\right)\qquad(i=0,1,2,\cdots) \tag{8.61}$$

则 $\widetilde{\mathrm{P}}_i(z)(i=0,1,2,\cdots)$ 是在 $[a,b]$ 上的正交多项式组，且满足如下正交条件

$$\int_a^b \widetilde{\mathrm{P}}_n(z)\widetilde{\mathrm{P}}_m(z)\mathrm{d}z = \begin{cases} \dfrac{b-a}{2n+1}, & m=n=0,1,2,\cdots \\ 0, & m\neq n \end{cases} \tag{8.62}$$

今取

$$Q_m(z) = \sum_{i=0}^{m} c_i \widetilde{\mathrm{P}}_i(z), \qquad a\leqslant z\leqslant b \tag{8.63}$$

与式(8.51)和式(8.55)相应的公式为

$$I_m = \int_a^b [Q_m(z) - f(z)]^2 \mathrm{d}z = \min \tag{8.64}$$

$$c_i = \frac{2i+1}{b-a}\int_a^b \widetilde{\mathrm{P}}_i(z) f(z)\mathrm{d}z \tag{8.65}$$

**例 8.7**　设 $f(x)=|x|,-1\leqslant x\leqslant 1$，求它的最小平方逼近多项式

$$Q_5(x) = \sum_{i=0}^{5} c_i \mathrm{P}_i(x) \tag{8.66}$$

**解**　按式(8.65)及勒让德多项式的性质得

$$c_1 = c_3 = c_5 = 0$$

$$c_{2k} = (4k+1)\int_0^1 x\mathrm{P}_{2k}(x)\mathrm{d}x \qquad (k=0,1,2)$$

所以

$$c_0 = \int_0^1 x\mathrm{d}x = \frac{1}{2}$$

$$c_2 = \frac{5}{2}\int_0^1 x(3x^2-1)\mathrm{d}x = \frac{5}{8}$$

$$c_4 = \frac{9}{8}\int_0^1 x(35x^4-30x^2+3)\mathrm{d}x = -\frac{3}{16}$$

将以上数值代入式(8.66)后得

$$\begin{aligned} Q_5(x) &= \frac{1}{2} + \frac{5}{16}(3x^2-1) - \frac{3}{128}(35x^4-30x^2+3) \\ &= \frac{15}{128}(-7x^4+14x^2+1) \end{aligned}$$

对于多变量函数在连续区间上的最小平方逼近问题可以仿照以上方法进行处理，这里不再叙述。

## §4　切比雪夫多项式及函数按切比雪夫多项式的展开式

切比雪夫多项式是函数逼近中的重要工具，下面具体介绍它的定义及有关性质。

### 4.1　切比雪夫多项式的定义

**定义 8.3**　$n$ 次切比雪夫多项式定义为

$$T_n(x) = \cos n\theta \quad (0\leqslant\theta\leqslant\pi;\quad n=0,1,2,\cdots) \tag{8.67}$$

通过变换

$$x=\cos\theta \quad (0\leqslant\theta\leqslant\pi;-1\leqslant x\leqslant 1) \tag{8.68}$$

可将(8.67)式的 $T_n(x)$ 化为 $x$ 的 $n$ 次多项式。由公式

$$\cos(n+1)\theta+\cos(n-1)\theta=2\cos\theta\cos n\theta \tag{8.69}$$

立即可得

$$T_{n+1}(x)=2xT_n(x)-T_{n-1}(x) \qquad (n=1,2,\cdots) \tag{8.70}$$

其中 $T_0(x)=\cos 0=1, T_1(x)=\cos\theta=x$。利用(8.70)可以逐次写出切比雪夫多项式组：

$$\begin{cases} T_0(x)=1 \\ T_1(x)=x \\ T_2(x)=2x^2-1 \\ T_3(x)=4x^3-3x \\ T_4(x)=8x^4-8x^2-1 \\ T_5(x)=16x^5-20x^3+5x \\ T_6(x)=32x^6-48x^4+18x^2-1 \end{cases} \tag{8.71}$$

等等。

## 4.2 切比雪夫多项式的一些性质

切比雪夫多项式除有上面的递推性质外，还有下面的一些重要性质。

### 4.2.1 $T_n(x)$的奇偶性

因 $x=\cos\theta$，所以

$$\theta=\arccos x$$

$$T_n(x)=\cos n\arccos x$$

$$T_n(-x)=\cos n\arccos(-x)=\cos n(\theta+\pi)=(-1)^nT_n(x) \tag{8.72}$$

由式(8.72)可见，当 $n$ 为奇数时，$T_n(-x)=-T_n(x)$，即 $T_n(x)$ 为奇函数；当 $n$ 为偶数时，$T_n(-x)=T_n(x)$，即 $T_n(x)$ 为偶函数。

### 4.2.2 $T_r(x)$的极值点与零点

由 $T_r(x)=\cos r\theta(0\leqslant\theta\leqslant\pi)$ 得表 8.8。

表 8.8

| $r$ | 极值点 $\theta_j$ | 零点 $\theta_k$ |
|---|---|---|
| 1 | $0\cdot\frac{\pi}{2}$, $2\cdot\frac{\pi}{2}$ | $1\cdot\frac{\pi}{2}$ |
| 2 | $0\cdot\frac{\pi}{4}$, $2\cdot\frac{\pi}{4}$, $4\cdot\frac{\pi}{4}$ | $1\cdot\frac{\pi}{4}$, $3\cdot\frac{\pi}{4}$ |
| 3 | $0\cdot\frac{\pi}{6}$, $2\cdot\frac{\pi}{6}$, $4\cdot\frac{\pi}{6}$, $6\cdot\frac{\pi}{6}$ | $1\cdot\frac{\pi}{6}$, $3\cdot\frac{\pi}{6}$, $5\cdot\frac{\pi}{6}$ |
| 4 | $0\cdot\frac{\pi}{8}$, $2\cdot\frac{\pi}{8}$, $4\cdot\frac{\pi}{8}$, $6\cdot\frac{\pi}{8}$, $8\cdot\frac{\pi}{8}$ | $1\cdot\frac{\pi}{8}$, $3\cdot\frac{\pi}{8}$, $5\cdot\frac{\pi}{8}$, $7\cdot\frac{\pi}{8}$ |

不难归纳得 $T_r(x)$ 的零点为

$$\theta_k:\ 1\cdot\frac{\pi}{2r},\ 3\cdot\frac{\pi}{2r},\ \cdots\ ,(2r-1)\cdot\frac{\pi}{2r}$$

或 $$\theta_k=(2k-1)\frac{\pi}{2r} \qquad (k=1,2,\cdots,r) \tag{8.73}$$

则 $$x_k=\cos\theta_k=\cos\frac{2k-1}{2r}\pi \qquad (k=1,2,\cdots,r) \tag{8.74}$$

或 $$x_k=\cos\frac{2(k+1)-1}{2r}\pi \qquad (k=0,1,2,\cdots,r-1) \tag{8.75}$$

为使 $x_k$ 值随 $k$ 的增加而增长,可令

$$\begin{aligned} x_k &= -\cos\frac{2(k+1)-1}{2r}\pi \\ &= -\cos\left\{-\left[\frac{2(k+1)-1}{2r}\pi\right]\right\} \\ &= \cos\left[\pi-\frac{2(k+1)-1}{2r}\pi\right] \\ &= \cos\frac{2(r-k)-1}{2r}\pi \qquad (k=0,1,2,\cdots,r-1) \end{aligned} \tag{8.76}$$

$T_r(x)$的极值点为

$$\theta_j:\ 0\cdot\frac{\pi}{2r},\quad 2\cdot\frac{\pi}{2r},\quad \cdots,\quad 2r\cdot\frac{\pi}{2r}$$

或 $$\theta_j=2j\cdot\frac{\pi}{2r}=\frac{j\pi}{r} \qquad (j=0,1,2,\cdots,r) \tag{8.77}$$

则 $$x_j=\cos\theta_j=\cos\frac{j}{r}\pi \qquad (j=0,1,2,\cdots,r) \tag{8.78}$$

同法,为使增长具有一致性,可令

$$x_j=-\cos\frac{j}{r}\pi=\cos(\frac{j}{r}\pi+\pi)=\cos\frac{r+j}{r}\pi \qquad (j=0,1,2,\cdots,r) \tag{8.79}$$

或 $$x_j=\cos\frac{r+j-1}{r}\pi \qquad (j=1,2,\cdots,r+1) \tag{8.80}$$

或 $$x_j=-\cos\frac{j}{r}\pi=-\cos(-\frac{j}{r}\pi)=\cos(\pi-\frac{j}{r}\pi)=\cos\frac{r-j}{r}\pi \qquad (j=0,1,2,\cdots,r) \tag{8.81}$$

### 4.2.3　正交性

**定理 8.2**　任何两个相异的切比雪夫多项式 $T_r(x)$及 $T_s(x)$对于权函数

$$\rho(x)=\frac{1}{\sqrt{1-x^2}}$$

在区间[−1,1]上都是互相正交的,并且

$$\int_{-1}^{+1}\frac{T_r(x)T_s(x)}{\sqrt{1-x^2}}\mathrm{d}x=\begin{cases}\pi, & r=s=0\\ \frac{\pi}{2}, & r=s\neq 0\\ 0, & r\neq s\end{cases} \tag{8.82}$$

**证**　因 $x=\cos\theta$,代入下式得

$$\begin{aligned}\int_{-1}^{+1}\frac{T_r(x)T_s(x)}{\sqrt{1-x^2}}\mathrm{d}x &= \int_0^{\pi}\cos r\theta\cos s\theta\,\mathrm{d}\theta \\ &= \frac{1}{2}\int_0^{\pi}[\cos(r-s)\theta+\cos(r+s)\theta]\mathrm{d}\theta\end{aligned}$$

$$= \frac{1}{2}\left[\frac{\sin(r-s)\theta}{r-s} + \frac{\sin(r+s)\theta}{r+s}\right]\Bigg|_0^\pi$$

$$= \frac{1}{2}\left[\frac{\sin(r-s)\pi}{r-s} + \frac{\sin(r+s)\pi}{r+s}\right] = 0 \quad (r \neq s)$$

当 $r=s=0$ 时，

$$\int_{-1}^{+1} \frac{T_0(x)T_0(x)}{\sqrt{1-x^2}}\mathrm{d}x = \int_\pi^0 \frac{1}{\sqrt{1-\cos^2\theta}}(-\sin\theta\mathrm{d}\theta) = \int_0^\pi \mathrm{d}\theta = \pi$$

当 $r = s \neq 0$ 时，

$$\int_{-1}^{+1} \frac{T_r(x)T_r(x)}{\sqrt{1-x^2}}\mathrm{d}x = \int_0^\pi \cos^2 r\theta\mathrm{d}\theta$$

$$= \frac{1}{2}\int_0^\pi (1+\cos 2r\theta)\mathrm{d}\theta$$

$$= \frac{1}{2}\left(\theta + \frac{\sin 2r\theta}{2r}\right)\Bigg|_0^\pi = \frac{\pi}{2}$$

定理得证。

此外，$T_r(x)$ 在 $T_{n+1}(x)$ 的零点集上亦具有正交性。

**定理 8.3** 任意两个相异的切比雪夫多项式 $T_r(x)$ 及 $T_s(x)(r,s \leqslant n)$ 在 $T_{n+1}(x)$ 的零点集

$$x_i = \cos\frac{2i-1}{2(n+1)}\pi \quad (i=1,2,\cdots,n+1) \tag{8.83}$$

上都是互相正交的，并且

$$\sum_{i=1}^{n+1} T_r(x_i)T_s(x_i) = \begin{cases} n+1, & r=s=0 \\ \dfrac{n+1}{2}, & 1 \leqslant r=s \leqslant n \\ 0, & r \neq s \end{cases} \tag{8.84}$$

**证** 利用余弦的和积公式

$$\cos r\theta \cdot \cos\theta = \frac{1}{2}[\cos(r+s)\theta + \cos(r-s)\theta]$$

简化下式

$$\sum_{i=1}^{n+1} T_r(x_i)T_s(x_i) = \sum_{i=1}^{n+1}\frac{1}{2}[\cos(r+s)\theta_i + \cos(r-s)\theta_i]$$

$$= \frac{1}{2}\sum_{i=1}^{n+1}\left[\cos(r+s)\frac{2i-1}{2(n+1)}\pi + \cos(r-s)\frac{2i-1}{2(n+1)}\pi\right]$$

$$= \frac{1}{2}\sum_{i=1}^{n+1}\left\{\cos\left[\frac{(r+s)\pi}{n+1}i - \frac{(r+s)\pi}{2(n+1)}\right] + \cos\left[\frac{(r-s)\pi}{n+1}i - \frac{(r-s)\pi}{2(n+1)}\right]\right\} \tag{8.85}$$

又

$$\sum_{i=1}^{N} 2\cos(hi+a)\sin\frac{h}{2}$$

$$= \sum_{i=1}^{N}\left\{\sin\left[\left(i+\frac{1}{2}\right)h+a\right] - \sin\left[\left(i-\frac{1}{2}\right)h+a\right]\right\}$$

$$= \sin\left[\left(N+\frac{1}{2}\right)h+a\right] - \sin\left(\frac{h}{2}+a\right)$$

所以
$$\sum_{i=1}^{N}\cos(hi+a)=\frac{\sin\left[(N+\frac{1}{2})h+a\right]-\sin(\frac{1}{2}h+a)}{2\sin\frac{h}{2}} \tag{8.86}$$

利用上式可把式(8.85)化为

$$\begin{aligned}
&\sum_{i=1}^{n+1}T_r(x_i)T_s(x_i)\\
&=\frac{1}{2}\left\{\frac{\sin\left\{\left[(n+1)+\frac{1}{2}\right]\frac{(r+s)\pi}{n+1}-\frac{(r+s)\pi}{2(n+1)}\right\}-\sin\left[\frac{(r+s)\pi}{2(n+1)}-\frac{(r+s)\pi}{2(n+1)}\right]}{2\sin\left[\frac{(r+s)\pi}{2(n+1)}\right]}+\right.\\
&\left.\frac{\sin\left\{\left[(n+1)+\frac{1}{2}\right]\frac{(r-s)\pi}{n+1}-\frac{(r-s)\pi}{2(n+1)}\right\}-\sin\left[\frac{(r-s)\pi}{2(n+1)}-\frac{(r-s)\pi}{2(n+1)}\right]}{2\sin\left[\frac{(r-s)\pi}{2(n+1)}\right]}\right\}\\
&=\frac{1}{2}\left\{\frac{\sin(r+s)\pi-\sin 0}{2\sin\frac{(r+s)\pi}{2(n+1)}}+\frac{\sin(r-s)\pi-\sin 0}{2\sin\frac{(r-s)\pi}{2(n+1)}}\right\}=0\quad(当\ r\neq s\ 时)
\end{aligned}$$

当 $r=s=0$ 时，

$$\sum_{i=1}^{n+1}T_0(x_i)T_0(x_i)=\sum_{i=1}^{n+1}1=n+1$$

当 $r=s\neq0$ 时，

$$\begin{aligned}
&\sum_{i=1}^{n+1}T_r(x_i)T_r(x_i)=\sum_{i=1}^{n+1}T_r^2(x_i)=\sum_{i=1}^{n+1}\cos^2 r\theta_i\\
&=\frac{1}{2}\sum_{i=1}^{n+1}[1+\cos 2r\theta_i]\\
&=\frac{1}{2}\sum_{i=1}^{n+1}1+\frac{1}{2}\sum_{i=1}^{n+1}\cos 2r\frac{(2i-1)\pi}{2(n+1)}\\
&=\frac{n+1}{2}+\frac{1}{2}\sum_{i=1}^{n+1}\cos\left[\frac{2r\pi}{n+1}i-\frac{r\pi}{n+1}\right]\\
&=\frac{n+1}{2}+\frac{1}{2}\,\frac{\sin\left\{\left[(n+1)+\frac{1}{2}\right]\frac{2r\pi}{n+1}-\frac{r\pi}{n+1}\right\}-\sin\left\{\frac{2r\pi}{2(n+1)}-\frac{r\pi}{n+1}\right\}}{2\sin\frac{2r\pi}{2(n+1)}}\\
&=\frac{n+1}{2}+\frac{1}{2}\,\frac{\sin 2r\pi-\sin 0}{2\sin\frac{r\pi}{n+1}}=\frac{n+1}{2}
\end{aligned}$$

定理得证。

### 4.3　函数按切比雪夫多项式的展开式

如果采用切比雪夫多项式构成正交多项式组，建立 $f(x)$在$[-1,+1]$上的最小平方逼近多项式

$$Q_m(x)=C_0T_0(x)+C_1T_1(x)+\cdots+C_mT_m(x) \tag{8.87}$$

则称 $Q_m(x)$ 为函数 $f(x)$ 按切比雪夫多项式的展开式。

由于切比雪夫多项式组在 $[-1,+1]$ 上是关于权函数 $\rho(x)=\frac{1}{\sqrt{1-x^2}}$ 的正交多项式组，因此在计算式(8.51)和式(8.55)中的积分号内均需乘以 $\rho(x)$ 得

$$I_m=\int_{-1}^{+1}\rho(x)[Q_m(x)-f(x)]^2\,\mathrm{d}x=\min \tag{8.88}$$

$$C_i=\frac{\int_{-1}^{+1}\rho(x)T_i(x)f(x)\mathrm{d}x}{\int_{-1}^{+1}\rho(x)T_i^2(x)\mathrm{d}x}=\begin{cases}\frac{1}{\pi}\int_{-1}^{+1}\rho(x)T_0(x)f(x)\mathrm{d}x, & i=0\\ \frac{2}{\pi}\int_{-1}^{+1}\rho(x)T_i(x)f(x)\mathrm{d}x, & i\neq 0\end{cases} \tag{8.89}$$

若将式(8.87)改写成

$$Q_m(x)=\frac{1}{2}C_0T_0(x)+C_1T_1(x)+\cdots+C_mT_m(x) \tag{8.90}$$

则式(8.89)中右端两式可统一为

$$C_i=\frac{2}{\pi}\int_{-1}^{+1}\rho(x)T_i(x)f(x)\mathrm{d}x \qquad (i=0,1,2,\cdots,m) \tag{8.91}$$

系数 $C_i$ 亦可用数值积分近似计算。由于 $x=\cos\theta,\mathrm{d}x=-\sin\theta\mathrm{d}\theta$，代入(8.91)式得

$$C_i=\frac{2}{\pi}\int_0^{\pi}\cos i\theta f(\cos\theta)\mathrm{d}\theta=\frac{2}{\pi}\int_0^{\pi}F(\theta)\mathrm{d}\theta \tag{8.92}$$

其中

$$F(\theta)=\cos i\theta f(\cos\theta)=T_i(x)f(x) \tag{8.93}$$

今将 $[0,\pi]$ 区间 $n$ 等分得

$$\Delta\theta=\frac{\pi-0}{n}=\frac{\pi}{n}$$

$$\theta_j=j\Delta\theta=\frac{j\pi}{n} \qquad (j=0,1,2,\cdots,n)$$

$$x_j=\cos\theta_j=\cos\frac{j\pi}{n} \qquad (j=0,1,2,\cdots,n)$$

使用复合梯形公式得

$$\begin{aligned}C_i&=\frac{2}{\pi}\cdot\frac{\pi}{n}\left[\frac{F(\theta_0)}{2}+F(\theta_1)+\cdots+F(\theta_{n-1})+\frac{F(\theta_n)}{2}\right]\\&=\frac{2}{n}\left[\frac{1}{2}f(x_0)T_i(x_0)+f(x_1)T_i(x_1)+\cdots+f(x_{n-1})T_i(x_{n-1})+\right.\\&\left.\frac{1}{2}f(x_n)T_i(x_n)\right] \qquad (i=0,1,2,\cdots,m)\end{aligned} \tag{8.94}$$

如果采用 $T_{n+1}(x)$ 的零点

$$x_i=\cos\frac{2i-1}{2(n+1)}\pi \qquad (i=1,2,\cdots,n+1)$$

作为离散点集，并利用正交性式(8.84)，则系数 $C_i$ 按式(8.37)计算得

$$C_i=\frac{\sum_{j=1}^{n+1}f(x_j)T_i(x_j)}{\sum_{j=1}^{n+1}T_i^2(x_j)}=\frac{\sum_{j=1}^{n+1}f(x_j)T_i(x_j)}{\frac{n+1}{2}}=\frac{2}{n+1}\sum_{j=1}^{n+1}f(x_j)T_i(x_j)$$

$$(i=0,1,2,\cdots,m\leqslant n) \tag{8.95}$$

当 $f(x)$ 为区间 $[a,b]$ 上的函数时，可作如下变换

$$t=\frac{b-a}{2}x+\frac{b+a}{2} \tag{8.96}$$

将区间$[a,b]$化成$[-1,+1]$的情况再进行处理。

**例 8.8**　使用 $T_4(x)$的零点集，求

$$f(x)=\frac{1}{1+x^2}$$

在$[-1,+1]$区间上的三次切比雪夫多项式的展开式。

**解**　取 $T_4(x)$的零点集为

$$x_j=\cos\frac{2j-1}{8}\pi \qquad (j=1,2,3,4)$$

按公式计算得

$$Q_3(x)=\frac{1}{2}C_0+C_1T_1(x)+C_2T_2(x)+C_3T_3(x)$$

$$C_0=\frac{2}{4}\sum_{j=1}^{4}f(x_j)T_0(x_j)=\frac{2}{4}\sum_{j=1}^{4}\frac{1}{1+x_j^2}=1.411\ 765$$

$$C_1=\frac{2}{4}\sum_{j=1}^{4}f(x_j)T_1(x_j)=\frac{2}{4}\sum_{j=1}^{4}\frac{x_j}{1+x_j^2}=0$$

$$C_2=\frac{2}{4}\sum_{j=1}^{4}f(x_j)T_2(x_j)=\frac{2}{4}\sum_{j=1}^{4}\frac{2x_j^2-1}{1+x_j^2}=-0.235\ 294$$

$$C_3=\frac{2}{4}\sum_{j=1}^{4}f(x_j)T_3(x_j)=\frac{2}{4}\sum_{j=1}^{4}\frac{4x_j^3-3x_j}{1+x_j^2}=0$$

将 $C_0$、$C_1$、$C_2$、$C_3$ 代入 $Q_3(x)$得

$$Q_3(x)=0.705\ 882-0.235\ 294T_2(x)=0.941\ 176-0.470\ 588x^2$$

函数按切比雪夫多项式的展开式有如下收敛性定理。

**定理 8.4**　如果 $f(x)$在$[-1,+1]$上连续，其一阶导数存在且有界，那么由式(8.90)、式(8.91)所定义的最小平方逼近多项式 $Q_m(x)$当 $m\to\infty$时一致收敛于 $f(x)$。(证明略)

在收敛情况下，设

$$f(x)=\frac{1}{2}C_0T_0(x)+\sum_{j=1}^{\infty}C_jT_j(x) \tag{8.97}$$

则近似式

$$Q_m(x)=\frac{1}{2}C_0T_0(x)+C_1T_1(x)+\cdots+C_mT_m(x)$$

相对于 $f(x)$的截断误差为

$$R_m(x)=f(x)-Q_m(x)=\sum_{j=m+1}^{\infty}C_jT_j(x)$$

如果 $C_{m+1}\neq 0$，且系数迅速收敛于 0，则有以下近似估计

$$|R_m(x)|\approx|C_{m+1}T_{m+1}(x)|\leqslant|C_{m+1}| \tag{8.98}$$

## §5　最佳一致逼近

采用多项式逼近函数时，从理论上讲，对于给定的 $f(x)\in C[a,b]$和给定的逼近精度$\varepsilon>0$，

能否找到一个代数多项式 $P_n(x)=a_0+a_1x+\cdots+a_nx^n$，使得

$$\| f(x)-P_n(x) \|_\infty \leqslant \varepsilon \tag{8.99}$$

成立。

对于这个问题，早在1885年魏尔斯特拉斯(Weierstrass)就给出了肯定的回答。

**定理8.5** 设 $f(x)\in C[a,b]$，则对于任意给定的 $\varepsilon>0$，都存在多项式 $P_n(x)$，使得

$$\| f(x)-P_n(x) \|_\infty \leqslant \varepsilon$$

成立。

这个定理有许多不同的证明方法，其中伯恩斯坦给出的证明理论上比较完美，它不仅证明了 $P_n(x)$ 的存在性，同时也可以用来构造 $P_n(x)$。他引进了如下的伯恩斯坦多项式

$$B_n(f)=\sum_{i=0}^{n} f\left(\frac{i}{n}\right)c_i^n x^i(1-x)^{n-i} \qquad (n=1,2,\cdots) \tag{8.100}$$

其中 $f(x)\in C[0,1]$，并证明了如下关系式

$$\lim_{n\to\infty} B_n(f)=f(x), \qquad x\in[0,1] \tag{8.101}$$

如果 $f(x)$ 在 $[0,1]$ 上的各阶导数存在，还可进一步证明 $B_n(f)$ 的各阶导数也分别收敛到 $f(x)$ 的相应阶导数。

尽管如此，直接利用(8.100)式去做数值逼近时，效果并不理想，最主要的问题是收敛太慢。因此为达到给定的精度要求，$n$ 的值就较大，又随着 $n$ 的增大，对 $B_n(f)$ 的计算会出现数值不稳定现象。这说明，用伯恩斯坦多项式构造近似函数 $P_n(x)$，只能在次数较低和精度要求不高的情况下才有效。

因此，很自然地提出下述问题，能否在所有次数不超过 $n$ 次的代数多项式集合 $H_n$ 上，找到一个在 $\| f(x)-P_n(x) \|_\infty$ 意义下最小的逼近多项式，这就导致了最佳一致逼近问题，这个问题也称为极大值取极小的问题。

### 5.1 最佳一致逼近的概念

设 $f(x)\in C[a,b]$，对于 $P_n(x)\in H_n$，称

$$\Delta(x)=f(x)-P_n(x) \tag{8.102}$$

为误差函数。$\Delta(x)$ 在 $[a,b]$ 上的最大绝对值

$$E=\| \Delta(x) \|_\infty=\max_{x\in[a,b]} |\Delta(x)| \tag{8.103}$$

称为偏差。对不同的 $P_k(x)$ $(k\leqslant n)$，其偏差值亦各不相同，其中必有最小值 $E_n(f)$

$$E_n(f)=\min_{P_n\in H_n} E=\min_{P_n\in H_n} \| \Delta(x) \|_\infty=\min_{P_n\in H_n} \{ \max_{x\in[a,b]} |f(x)-P_n(x)| \} \tag{8.104}$$

称 $E_n(f)$ 为 $n$ 次最小偏差或 $n$ 次最佳逼近值。并称满足式(8.104)的多项式 $P_n^*(x)$ 为 $f(x)$ 在 $[a,b]$ 上的最佳一致逼近多项式或最佳逼近多项式，$P_n^*(x)$ 满足

$$\max_{x\in[a,b]} |f(x)-P_n^*(x)|=\max_{x\in[a,b]} |\Delta^*(x)|=E_n(f) \tag{8.105}$$

我们把满足上式的那些 $x$ 值统称为偏差点，且依 $\Delta^*(x)$ 的符号的正、负不同称为正偏差点或负偏差点。

### 5.2 最佳一致逼近多项式的有关理论问题

针对上述定义的最佳一致逼近多项式，需要研究的问题有它是否存在且唯一？当存在唯一的情况下又如何寻找或构造它？对这些问题的回答构成了最佳一致逼近研究的中心内容。

关于最佳一致逼近多项式的存在性已由定理 8.5 做了回答。存在性的证明不是本书重点，而构造性的证明，迄今尚未完满地解决。但切比雪夫给出了描述最佳一致逼近多项式的重要特征定理，有助于最佳一致逼近多项式的求解，其唯一性也可由此推得。在叙述切比雪夫定理以前，先引入以下定义。

**定义 8.4**　对于 $f(x), P_n^*(x)\in C[a,b]$，如果它的偏差点集满足

$$\begin{cases} a\leqslant x_1<x_2<\cdots<x_k\leqslant b \\ \Delta(x_j)=-\Delta(x_{j+1}) \qquad (j=1,2,\cdots,k-1) \\ |\Delta(x_j)|=E_n(f) \qquad (j=1,2,\cdots,k) \end{cases} \tag{8.106}$$

则称上述点集为 $f(x)$ 在 $[a,b]$ 上的一个交错点组。

**例 8.9**　设 $f(x)=\sin x, x\in[0,\pi]$，求 $P_0^*(x)$ 及交错点组。

**解**　设 $P_0^*(x)=C$，建立误差函数

$$\Delta(x)=\sin x-C$$

调节 $C$ 值，使 $\Delta(x)$ 在 $[0,\pi]$ 上的最大绝对值达到最小，显见 $C=0.5$（图 8.1），则 $P_0^*(x)=0.5$，

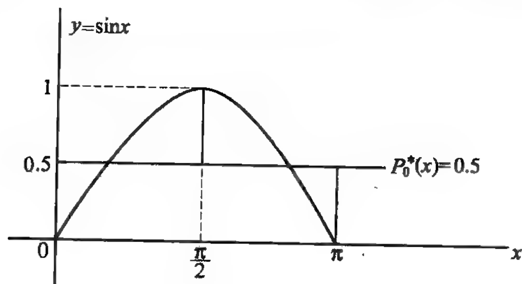



图 8.1

$$E_0(f)=\max_{x\in[0,\pi]}|\Delta^*(x)|=\max_{x\in[0,\pi]}|\sin x-0.5|=0.5$$

且 $|\Delta^*(x)|$ 在 $x=0,\frac{\pi}{2},\pi$ 上取得 $E_0(f)$ 值，相应的 $\Delta^*(x)$ 的符号为 $-,+,-$，所以 $f(x)=\sin x$ 在 $[0,\pi]$ 上的交错点组为 $0,\frac{\pi}{2},\pi$。一般情况下，交错点组往往是不唯一的，如图 8.2 所示。图中 $(x_1,x_2,x_3)$ 是交错点组，$(x_1,\tilde{x}_2,x_3)$ 也是。

有了以上准备，就可以叙述关于 $C[a,b]$ 上 $f(x)$ 的最佳一致逼近多项式的特征定理了。

**定理 8.6**　（交错点组定理）设 $f(x)\in C[a,b]$，$P_n^*(x)$ 为一个次数不超过 $n$ 的最佳一致逼近多项式，则 $\Delta^*(x)=f(x)-P_n^*(x)$ 至少存在由 $n+2$ 个偏差点构成的交错点组（称为切比雪夫交错点组）。

**证**　设 $P_n^*(x)$ 是 $f(x)$ 在 $[a,b]$ 上的最佳一致逼近多项式，由于 $\Delta^*(x)=f(x)-P_n^*(x)$ 在闭区间上连续，则 $\Delta^*(x)$ 在 $[a,b]$ 上必定一致连续[❶]。因此总可以把区间 $[a,b]$ 分成若干个子区间，使 $\Delta^*(x)$ 在每个子区间 $l$ 上的振幅小于 $\frac{1}{2}E_n(f)$，现把上述子区间再分成两类，一类为

❶　一致连续指的是对任意取定的 $\varepsilon>0$，总存在有 $\delta>0$，对 $[a,b]$ 中的任意 $x',x''$，只要 $|x'-x''|<\delta$，就有 $|f(x')-f(x'')|<\varepsilon$ 成立。

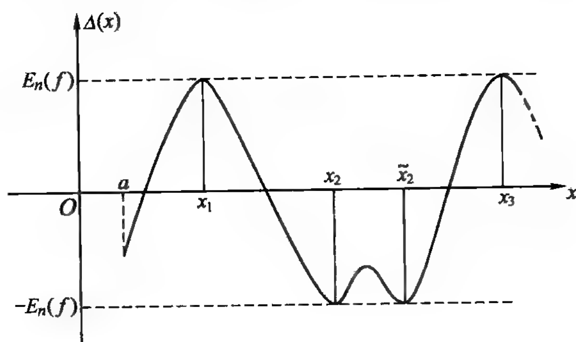



图 8.2

包含有$\Delta^*(x)$的偏差点的子区间，从左到右记作$I_1$、$I_2$，…，$I_{N1}$。其余的子区间归为第二类，记作$J_1$，$J_2$，…，$J_{N2}$。从上述子区间的做法知

$$|\Delta^*(x)|=|f(x)-P_n^*(x)|>\frac{1}{2}E_n(f),\quad x\in I_j\qquad (j=1,2,\cdots,N_1)\tag{8.107}$$

$$|\Delta^*(x)|=|f(x)-P_n^*(x)|<\frac{1}{2}E_n(f),\quad x\in J_j\qquad (j=1,2,\cdots,N_2)\tag{8.108}$$

由式(8.107)知，在每个$I_j$中，$\Delta^*(x)$的符号总是固定的。为确定起见，不妨设在$I_1$上$\Delta^*(x)$的符号是正的。对于第一类中的子区间继续按$\Delta^*(x)$符号的+、-，从左到右地分成若干组

$$I_1,I_{k_0+2},\cdots,I_{k_1}\qquad (\Delta^*(x)>0)$$

$$I_{k_1+1},I_{k_1+2},\cdots,I_{k_2}\qquad (\Delta^*(x)<0)$$

$$\cdots$$

$$I_{k_{m-1}+1},I_{k_{m-1}+2},\cdots,I_{k_s}\qquad (\Delta^*(x)\cdot(-1)^{s-1}>0)$$

显然这$s$组中的每个子区间依次包含有$\Delta^*(x)$的正偏差点与负偏差点。下面证明$s\geqslant n+2$，用反证法来证。假若不然，即$s<n+2$，由于$\Delta^*(x)$在$I_{k_1}$与$I_{k_1+1}$上不同号，所以$I_{k_1}$的右端点不能与$I_{k_1+1}$的左端点重合，于是就可在$I_{k_1}$与$I_{k_1+1}$之间任意取定一点$x_1$，它不属于上述$I_{k_1}$和$I_{k_1+1}$之一。类似地，在每个$I_{k_i}$与$I_{k_i+1}$之间可以任意取定一点$x_i(i=2,3,\cdots,s-1)$，由这些点建立以下两个多项式

$$\begin{cases}q(x)=(x_1-x)(x_2-x)\cdots(x_{s-1}-x)\\ Q(x)=P_n^*(x)+\varepsilon q(x)\end{cases}\tag{8.109}$$

式中，$\varepsilon$为待定参数。因$s\leqslant n+1$，所以$q(x)$为次数不超过$n$次的多项式，因知$Q(x)\in H_n$。考虑误差函数

$$\widetilde{\Delta}(x)=f(x)-Q(x)=f(x)-P_n^*(x)-\varepsilon q(x)=\Delta^*(x)-\varepsilon q(x)\tag{8.110}$$

今证只需适当选取充分小的$\varepsilon>0$，可使

$$\max_{a\leqslant x\leqslant b}|\widetilde{\Delta}(x)|<E_n(f)$$

成立，从而导致矛盾，定理得证。注意到$q(x)$在且仅在$x_i(i=1,2,\cdots,s-1)$处变号，又因$q(x)$在$I_1$上是正的，因而$q(x)$与$\Delta^*(x)$在$I_j(j=1,2,\cdots,N_1)$上同号。记

$$E^*=\max_{1\leqslant i\leqslant N_2}\max_{x\in J_i}|\Delta^*(x)|$$

由式(8.108)知 $E^* < \frac{1}{2}E_n(f)$。现选取 $\varepsilon>0$ 充分小，使下式

$$\varepsilon \max_{a\leqslant x\leqslant b}\left|q(x)\right| < \min\left\{E_n(f)-E^*, \frac{1}{2}E_n(f)\right\}$$

成立。则当 $x\in J_i(i=1,2,\cdots,N_2)$时有

$$\max_{x\in J_i}\left|\tilde{\Delta}(x)\right| \leqslant \max_{x\in J_i}\left|\Delta^*(x)\right| + \varepsilon\max_{x\in J_i}\left|q(x)\right|$$

$$< E^* + \min\left\{E_n(f)-E^*, \frac{1}{2}E_n(f)\right\} < E_n(f)$$

上式表明，与 $\Delta^*(x)$的峰值比较，$\tilde{\Delta}(x)$在 $J_i(i=1,2,\cdots,N_2)$上的峰值并未增加。而在 $I_j(j=1,2,\cdots,N_1)$上，由于 $\Delta^*(x)$与 $q(x)$同号，且 $q(x)\neq 0$，所以有

$$\max_{x\in I_j}\left|\tilde{\Delta}(x)\right| = \max_{x\in I_j}\left|f(x)-Q(x)\right| = \max_{x\in I_j}\left|\Delta^*(x)-\varepsilon q(x)\right|$$

$$< \max_{x\in I_j}\left|\Delta^*(x)\right| = E_n(f)$$

由此可见，$Q(x)$是比 $P_n^*(x)$更好的逼近，这就导致矛盾。在本定理的论述中，如果 $s\geqslant n+2$，我们将不能提供 $Q(x)\in H_n$，并在这种情况下导出矛盾。

**定理 8.7**　设 $f(x)\in C[a,b]$，又设 $P_n(x)\in H_n$，且至少具有$(n+2)$个点构成的切比雪夫交错点组，则 $P_n(x)$是 $f(x)$在$[a,b]$上的一个最佳一致逼近多项式。

**证**　因为在切比雪夫交错点组上

$$|f(x_k)-P_n(x_n)| = \max_{a\leqslant x\leqslant b}|f(x)-P_n(x)| = E_n(f)$$

其中 $x_k(k=1,2,\cdots,n+2)$为偏差点。今设 $P_n(x)$不是最佳一致逼近多项式，则有多项式 $Q_n(x)$存在，使

$$\max_{a\leqslant x\leqslant b}|f(x)-Q_n(x)| < E_n(f)$$

于是多项式

$$Q_n(x)-P_n(x) = [f(x)-P_n(x)]-[f(x)-Q_n(x)] \tag{8.111}$$

在偏差点 $x_k(k=1,2,\cdots,n+2)$上的符号与 $f(x)-P_n(x)$的符号一样，因而在上述$(n+2)$个点上依次取正负号，这样就推知 $Q_n(x)-P_n(x)$在$[a,b]$上至少有$(n+1)$个零点，但这是不可能的，因为 $Q_n(x)-P_n(x)$是不恒等于零且次数又不超过 $n$ 的多项式。(证毕)

上述定理称为切比雪夫定理。这个定理刻画了最佳一致逼近多项式特征，那就是 $\Delta^*(x)$在交错点组的$(n+2)$个点上取得相同的幅值 $E_n(f)$，且符号依次正负交替改变。因此可视 $\Delta^*(x)$是在$[a,b]$上作等幅度振荡的函数或称最佳一致逼近为误差函数作等幅度振荡的逼近，故而误差分布较均匀。最佳一致逼近多项式的上述特性，已成为求取最佳一致逼近多项式的各种近似解法的出发点和依据。这里还可以解决最佳一致逼近多项式的唯一性问题。

**定理 8.8**　设 $f(x)\in C[a,b]$，则在 $H_n$ 中的最佳一致逼近多项式是唯一的。

**证**　设在 $H_n$ 中存在有两个不同的最佳一致逼近多项式 $P_n(x)$和 $Q_n(x)$，即有

$$\max_{a\leqslant x\leqslant b}\left|f(x)-P_n(x)\right| = E_n(f),\quad \max_{a\leqslant x\leqslant b}\left|f(x)-Q_n(x)\right| = E_n(f) \tag{8.112}$$

令　$R_n(x)=\frac{1}{2}(P_n(x)+Q_n(x))$，因为

$$\max_{a\leqslant x\leqslant b}\left|f(x)-R_n(x)\right| = \max_{a\leqslant x\leqslant b}\left|f(x)-\frac{1}{2}(P_n(x)+Q_n(x))\right|$$

$$\leqslant \frac{1}{2}\max_{a\leqslant x\leqslant b}\left|f(x)-P_n(x)\right|+\frac{1}{2}\max_{a\leqslant x\leqslant b}\left|f(x)-Q_n(x)\right|$$

$$=\frac{1}{2}E_n(f)+\frac{1}{2}E_n(f)=E_n(f)$$

所以 $R_n(x)$ 也是最佳一致逼近多项式。由定理 8.6 知，$R_n(x)$ 至少有 $n+2$ 个正负偏差点 $x_1$，$x_2,\cdots,x_{n+2}$ 满足下式

$$E_n(f)=|f(x_i)-R_n(x_i)|=\left|\frac{f(x_i)-P_n(x_i)}{2}+\frac{f(x_i)-Q_n(x_i)}{2}\right|$$

$$(i=1,2,\cdots,n+2) \tag{8.113}$$

故 $f(x_i)-P_n(x_i)$ 和 $f(x_i)-Q_n(x_i)$ 必定同号，再据式(8.112)知，下式

$$f(x_i)-P_n(x_i)=f(x_i)-Q_n(x_i) \qquad (i=1,2,\cdots,n+2)$$

必定成立，即得

$$P_n(x_i)-Q_n(x_i)\equiv 0 \qquad (i=1,2,\cdots,n+2)$$

成立。由此推知 $P_n(x)-Q_n(x)$ 有 $n+1$ 个零点存在，这只有当 $P_n(x)\equiv Q_n(x)$ 时才可能。(证毕)

前已指出，切比雪夫交错点组往往是不唯一的，这会给数值计算带来一定的困难。然而在一定条件下，切比雪夫交错点组不仅唯一，而且切比雪夫交错点组中的个别点也可以被确定下来。

**定理 8.9** 如果函数 $f(x)$ 在 $[a,b]$ 上有 $f^{(n+1)}(x)\neq 0$（即 $f^{(n+1)}(x)$ 保号），则切比雪夫交错点组恰好由 $(n+2)$ 个偏差点构成，其中 $n$ 个偏差点为内偏差点，其余两个偏差点均是端点。

**证** 采用反证法。设 $a,b$ 都不是偏差点，则在 $(a,b)$ 内至少有 $(n+2)$ 个偏差点，这些偏差点也应视作误差函数的极值点，因此得到下式

$$[\Delta^*(x_i)]'=0 \qquad (i=1,2,\cdots,n+2)$$

逐次应用洛尔定理可得

$$[\Delta^*(\xi_i)]^{(n+1)}=f(\xi_i)^{(n+1)}-[P_n^*(\xi_i)]^{(n+1)}=f^{(n+1)}(\xi_i)=0 \qquad (i=1,2)$$

与定理条件不符，故假设不成立。

其次，设 $a,b$ 中之一不是偏差点，此时内偏差点至少有 $(n+1)$ 个，在这些点上，误差函数的一阶导数为 0，逐次应用洛尔定理可推知至少有一点 $\xi$，使 $[\Delta^*(\xi)]^{(n+1)}=f^{(n+1)}(\xi)-[P_n^*(\xi)]^{(n+1)}=f^{(n+1)}(\xi)=0$，仍与定理条件不符，由此证得 $a,b$ 必须都是偏差点不可。

最后证明内偏差点的个数只能是 $n$ 个，如若不然，则在 $(a,b)$ 内有多于 $n$ 个偏差点使误差函数的一阶导数为 0，仿上可推出至少存在一个 $\xi$，使 $f^{(n+1)}(\xi)=0$，与定理条件不符，从而证得在 $(a,b)$ 内只能有 $n$ 个内偏差点。

**定义 8.5** 若误差函数的切比雪夫交错点组恰好由 $n+2$ 个偏差点构成，且区间端点 $a,b$ 均是偏差点，则称该误差函数为标准的误差函数，其对应的图形称为标准的误差曲线；否则称为非标准的误差函数或非标准的误差曲线。

非标准的误差函数指其切比雪夫交错点组的偏差点个数多于 $(n+2)$ 个或端点之一或两者都不为偏差点的一类误差函数。例如，$f(x)=(x-1)^2$ 在 $[0,2]$ 上的最佳零次一致逼近多项式为 $P_0^*(x)=0.5$，其交错点个数至少为 $0+2=2$ 个，实际个数为 3 个：0,1,2。其误差函数属非标准误差函数。又如 $f(x)=\cos\frac{1}{4}\pi x$ 在 $[-1,+1]$ 上的七次最佳一致逼近多项式为

$$P_7^*(x)=0.999\,999\,972\,4-0.308\,424\,253\,6x^2+0.015\,849\,915\,3x^4-$$
$$0.000\,318\,880\,5x^6+0\cdot x^7 \tag{8.114}$$

因 $$f^{(8)}(x)=(\frac{\pi}{4})^8\cos\frac{1}{4}\pi x>0 \qquad (-1\leqslant x\leqslant +1)$$

所以 $$\Delta^*(x)=\cos\frac{1}{4}\pi x-P_7^*(x)$$

为标准的误差函数,其偏差点个数至少为 7+2=9 个,实际正好是 9 个,且 $\pm 1$ 均是偏差点。

在满足定理 8.9 中假设的条件下,最佳一致逼近多项式的求解问题就可归结为 $(a,b)$ 内 $n$ 个交错点组的求取问题。当 $n=1$ 时,只要在区间内求一个偏差点,问题就很简单了。

**例 8.10** 求 $e^x$ 在 $[-1,+1]$ 上的一次最佳一致逼近多项式 $P_1^*(x)=a_0^*+a_1^*x$。

**解** 因 $f(x)=e^x$,$f''(x)=e^x>0$(恒正),所以误差函数

$$\Delta(x)=e^x-(a_0^*+a_1^*x)$$

为标准的误差函数,故偏差点共有 3 个:$-1,x_1,+1$。因得以下等幅度振荡方程组

$$\begin{cases}\Delta(-1)=\Delta(+1)\\ \Delta(-1)=-\Delta(x_1)\\ \Delta'(x_1)=0\end{cases}$$

解得 $$\begin{cases}a_1^*=\dfrac{e-e^{-1}}{2}=1.175\ 2\\ x_1=\ln a_1^*=0.161\ 4\\ a_0^*=\dfrac{e^{-1}+e^{x_1}}{2}-a_1\dfrac{x_1-1}{2}=1.264\ 3\end{cases}$$

则 $$P_1^*(x)=1.264\ 3+1.175\ 2x$$

关于 $P_n^*(x)$ 的奇偶性有以下定理。

**定理 8.10** 若 $P_n^*(x)$ 为奇函数 $f(x)$ 在 $[-a,+a]$ 区间上的最佳一致逼近多项式,则 $P_n^*(x)$ 也是奇函数;若 $P_n^*(x)$ 为偶函数 $f(x)$ 在 $[-a,+a]$ 区间上的最佳一致逼近多项式,则 $P_n^*(x)$ 也是偶函数。

**证** 设 $f(x)$ 为区间 $[-a,+a]$ 上的偶函数,即 $f(x)=f(-x)$。因它的最佳一致逼近多项式 $P_n^*(x)$ 在 $[-a,0]$,$[0,a]$ 上的最大误差绝对值应相等,所以当 $x\in[-a,+a]$ 时,有下式成立

$$\begin{aligned}\|f(x)-P_n^*(x)\|_\infty &= \|f(-x)-P_n^*(-x)\|_\infty\\ &= \|f(x)-P_n^*(-x)\|_\infty \qquad (8.115)\end{aligned}$$

因在 $[-a,+a]$ 区间上 $f(x)$ 的最佳一致逼近多项式是唯一的,故得 $P_n^*(x)=P_n^*(-x)$,即 $P_n^*(x)$ 也是偶函数。

若 $f(x)$ 是奇函数,即 $f(x)=-f(-x)$,则当 $x\in[-a,+a]$ 时有下式成立

$$\begin{aligned}\|f(x)-P_n^*(x)\|_\infty &= \|f(-x)-P_n^*(-x)\|_\infty\\ &= \|-f(x)-P_n^*(-x)\|_\infty\\ &= \|f(x)+P_n^*(-x)\|_\infty\\ &= \|f(x)-(-P_n^*(-x))\|_\infty \qquad (8.116)\end{aligned}$$

由最佳一致逼近多项式的唯一性知 $P_n^*(x)=-P_n^*(-x)$,即 $P_n^*(x)$ 也是奇函数。

当 $P_n^*(x)$ 是偶函数时,则 $P_n^*(x)$ 的表达式中仅含有 $x$ 的偶次方项。设 $P_n^*(x)$ 的最高偶次方为 $2m$,这时应视 $P_n^*(x)$ 为 $2m+1$ 次的最佳一致逼近多项式

$$P_n^*(x)=P_{2m}(x)+0\cdot x^{2m+1}=P_{2m+1}^*(x) \qquad (8.117)$$

它的偏差点个数至少有$(2m+1)+2=2m+3$个，且$x=0$是偏差点。这时只需讨论$[0,a]$区间上的情况就行了。偶函数的例子如式(8.114)所示。

同样，当$P_n^*(x)$为奇函数时，$P_n^*(x)$的表达式中仅含有$x$的奇次方项。设$P_n^*(x)$的最高奇次方为$2k+1$，这时应视$P_n^*(x)$为$2k+2$次最佳一致逼近多项式

$$P_n^*(x)=P_{2k+1}(x)+0\cdot x^{2k+2}=P_{2k+2}^*(x) \tag{8.118}$$

它的偏差点个数至少有$(2k+2)+2=2k+4$个。奇函数的例子如$\arctan x$在$[-1,+1]$上的六次最佳一致逼近多项式

$$P_6^*(x)=0.995\,358x-0.288\,690x^3+0.079\,339x^5$$

其偏差点个数至少为$6+2=8$个，实际正好是8个，且$\pm1$亦是偏差点。

## 5.3 近似最佳一致逼近多项式

### 5.3.1 切比雪夫多项式的极性

1857年切比雪夫提出这样的问题：在$[-1,+1]$上最高次幂的系数为1的一切多项式中，求一个多项式使它在$[-1,+1]$上的最大绝对值为最小。下面的定理回答了这个问题。

**定理 8.11** 在所有最高次幂的系数为1的$n+1$次多项式中，在区间$[-1,+1]$上与零偏差最小的多项式是$\widetilde{T}_{n+1}(x)$。这里$\widetilde{T}_{n+1}(x)$是最高次幂的系数为1的切比雪夫多项式

$$\widetilde{T}_{n+1}(x)=\frac{1}{2^n}T_{n+1}(x) \tag{8.119}$$

**证** 首先建立误差函数

$$\Delta(x)=\widetilde{T}_{n+1}(x)-0=\frac{1}{2^n}T_{n+1}(x) \tag{8.120}$$

由于$T_{n+1}(x)$在$[-1,+1]$上具有极值点

$$x_j=\cos\frac{j}{n+1}\pi \qquad (j=0,1,2,\cdots,n+1)$$

代入式(8.120)得

$$\begin{aligned}\Delta(x_j)&=\frac{1}{2^n}\cos(n+1)\arccos x_j\\&=\frac{1}{2^n}\cos(n+1)\cdot\frac{j}{n+1}\pi\\&=\frac{1}{2^n}\cos j\pi=\frac{(-1)^j}{2^n}\quad (j=0,1,2,\cdots,n+1)\end{aligned} \tag{8.121}$$

现将式(8.120)改写为

$$\Delta(x)=\widetilde{T}_{n+1}(x)=x^{n+1}-P_n(x) \tag{8.122}$$

则式(8.121)表明，$P_n(x)$作为$f(x)=x^{n+1}$在$[-1,+1]$上的逼近多项式，其误差函数在$[-1,+1]$上具有符号相间的偏差点$(n+2)$个，据切比雪夫定理知，$P_n(x)$为$f(x)=x^{n+1}$在$[-1,+1]$上的最佳一致逼近多项式，即

$$\max_{-1\leqslant x\leqslant +1}|\Delta(x)|=\min_{P_n\in H_n}\|x^{n+1}-P_n(x)\|_\infty=E_n(f)$$

所以$\widetilde{T}_{n+1}(x)$是$[-1,+1]$上与0偏差最小的多项式。（证毕）

切比雪夫多项式的上述特性称为切比雪夫多项式的极性，这是一个很重要的性质，它表明以切比雪夫多项式为余式的误差幅值在整个区间$[-1,+1]$上是均匀分布的，因此这一性质在

求函数的近似最佳一致逼近多项式中有广泛的应用。下面介绍三种 $f(x)$在$[-1,+1]$上的近似最佳一致逼近多项式 $P_n(x)$的求法。

**5.3.2 截断的切比雪夫级数**

函数 $f(x)$在$[-1,+1]$上按切比雪夫多项式的展开式若取项数至 $T_n(x)$为止，则得

$$P_n(x)=\frac{1}{2}C_0T_0(x)+C_1T_1(x)+\cdots+C_nT_n(x) \tag{8.123}$$

当式中系数迅速收敛于 0 时，其与 $f(x)$的误差函数可近似为

$$\Delta(x)=f(x)-P_n(x)\approx C_{n+1}T_{n+1}(x) \tag{8.124}$$

上述误差函数在 $T_{n+1}(x)$的极值点

$$x_k=\cos\frac{k\pi}{n+1}\qquad (k=0,1,2,\cdots,n+1) \tag{8.125}$$

上取到相同的极大值$|C_{n+1}|$且符号依次正负相间，按切比雪夫定理知，上述 $P_n(x)$很接近于 $P_n^*(x)$，因此将它作为 $f(x)$在$[-1,+1]$上的近似最佳一致逼近多项式是很合适的。

**5.3.3 切比雪夫多项式零点插值法**

**定理 8.12** 如果 $f(x)\in C[a,b]$，则其 $n$ 次最佳一致逼近多项式就是 $f(x)$在$[a,b]$上的某个 $n$ 次插值多项式。

**证** 设 $P_n^*(x)$为 $f(x)$在$[a,b]$上的最佳一致逼近多项式，由切比雪夫定理知，其误差函数 $\Delta^*(x)=f(x)-P_n^*(x)$在$[a,b]$上至少有 $n+2$ 个正负相间的偏差点，这意味着 $\Delta^*(x)$在$[a,b]$上至少有 $n+1$ 个零点，于是 $P_n^*(x)$刚好是以这 $n+1$ 个零点为插值节点的插值多项式。（证毕）

上述定理表明，只要适当地选取插值节点，通过插值法就可获得最佳一致逼近多项式。设 $f(x)\in C[-1,+1]$，其在 $x_0,x_1,\cdots,x_n$ 上的插值多项式为 $P_n(x)$，则

$$\begin{aligned}R_n(x)=f(x)-P_n(x)&=\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)(x-x_1)\cdots(x-x_n)\\&=\frac{f^{(n+1)}(\xi)}{(n+1)!}\prod\nolimits_{n+1}(x)\end{aligned} \tag{8.126}$$

因$|f^{(n+1)}(x)|$在$[-1,+1]$上的最大值是确定的，所以要减小 $R_n(x)$，唯一的途径就是使多项式$\prod_{n+1}(x)$的最大绝对值为最小。因$\prod_{n+1}(x)$为最高次幂系数为 1 的多项式，因此可令

$$\prod\nolimits_{n+1}(x)=\widetilde{T}_{n+1}(x)=\frac{1}{2^n}T_{n+1}(x) \tag{8.127}$$

就能使$\prod_{n+1}(x)$具有最小偏差，这表示应取 $T_{n+1}(x)$的零点 $x_i(i=0,1,2,\cdots,n)$作为插值节点就能使

$$\max_{-1\leqslant x\leqslant 1}|(x-x_0)(x-x_1)\cdots(x-x_n)|=\max_{-1\leqslant x\leqslant 1}|\widetilde{T}_{n+1}(x)|$$

达到最小值 $1/2^n$，这时相应的余式有以下估计

$$\begin{aligned}|R_n(x)|\leqslant\left|\frac{M_{n+1}}{(n+1)!}\widetilde{T}_{n+1}(x)\right|&=\left|\frac{M_{n+1}}{2^n(n+1)!}T_{n+1}(x)\right|\\&=|C_{n+1}T_{n+1}(x)|\leqslant|C_{n+1}|\end{aligned} \tag{8.128}$$

注意，如果 $C_{n+1}=0$，则插值节点应在更高次的切比雪夫多项式的零点集上选取，通常是取 $T_{n+2}(x)$的零点集作为插值节点。比如 $f(x)$为奇函数或偶函数的情况就是这样。

若 $f(x)\in C[a,b]$，可作变换

$$x=\frac{b-a}{2}t+\frac{a+b}{2},\qquad x\in[a,b],\quad t\in[-1,+1] \tag{8.129}$$

$$\prod_{n+1}(x)=\left(\frac{b-a}{2}\right)^{n+1}(t-t_0)(t-t_1)\cdots(t-t_n)$$

取 $\widetilde{T}_{n+1}(t)$ 的零点

$$t_k=\cos\frac{2k+1}{2(n+1)}\pi\qquad(k=0,1,2,\cdots,n) \tag{8.130}$$

代入式(8.129)得插值节点

$$x_k=\frac{b-a}{2}\cos\frac{2k+1}{2(n+1)}\pi+\frac{a+b}{2}\qquad(k=0,1,2,\cdots,n) \tag{8.131}$$

相应地

$$\max_{a\leqslant x\leqslant b}\left|\prod_{n+1}(x)\right|\leqslant\left(\frac{b-a}{2}\right)^{n+1}\cdot\frac{1}{2^n} \tag{8.132}$$

$$\left|R_n(x)\right|\leqslant\frac{M_{n+1}}{(n+1)!}\left(\frac{b-a}{2}\right)^{n+1}|\widetilde{T}_{n+1}(t)|=C_{n+1}|\widetilde{T}_{n+1}(t)| \tag{8.133}$$

按式(8.131)的插值节点建立的插值多项式可作为 $f(x)$ 在 $[a,b]$ 上的近似最佳一致逼近多项式。

### 5.3.4　缩短幂级数法

一般情况下，函数 $f(x)$ 在 $[-1,+1]$ 上的台劳级数是容易得到的，当 $n$ 很大时，偏差

$$\max_{-1\leqslant x\leqslant 1}|f(x)-P_n(x)|$$

也能很小，但误差分布极不均匀，且方次较高。为改善它，可利用切比雪夫多项式的极性，对台劳级数的部分和进行改造，将它的高次幂减缩下来，同时使误差分布更加均匀，方法如下。

对指定的误差限 $\varepsilon$，先对函数进行台劳展开得

$$P_n(x)=a_0+a_1x+\cdots+a_nx^n=ax^n+P_{n-1}(x) \tag{8.134}$$

其展开的次数可以高一些，项数就相应地多一些，使余式的误差满足

$$|R_n(x)|\leqslant\varepsilon_n<\varepsilon \tag{8.135}$$

以后使用切比雪夫多项式将 $P_n(x)$ 的方次降低一次得到 $P_{n(n-1)}(x)$，计算公式设计如下。令

$$\begin{aligned}\Delta_{n-1}(x)&=P_n(x)-P_{n(n-1)}(x)\\&=a_nx^n+P_{n-1}(x)-P_{n(n-1)}(x)\\&=a_n\left[x^n+\frac{P_{n-1}(x)-P_{n(n-1)}(x)}{a_n}\right]\end{aligned} \tag{8.136}$$

上式方括号中是一个最高次幂系数为 1 的 $n$ 次多项式，若取

$$x^n+\frac{P_{n-1}(x)-P_{n(n-1)}(x)}{a_n}=\widetilde{T}_n(x)=\frac{1}{2^{n-1}}T_n(x) \tag{8.137}$$

时，$\|\Delta_{n-1}(x)\|_\infty=\min$。由(8.137)式可求出

$$P_{n(n-1)}(x)=P_n(x)-\frac{a_n}{2^{n-1}}T_n(x) \tag{8.138}$$

按上式求出的 $P_{n(n-1)}(x)$ 就是 $P_n(x)$ 在 $[-1,+1]$ 上的 $n-1$ 次最佳一致逼近多项式，其间的误差为

$$\|P_n(x)-P_{n(n-1)}(x)\|_\infty=\frac{|a_n|}{2^{n-1}}=\varepsilon_{n-1} \tag{8.139}$$

因　$\|f(x)-P_{n(n-1)}(x)\|_\infty\leqslant\|f(x)-P_n(x)\|_\infty+\|P_n(x)-P_{n(n-1)}(x)\|_\infty$

$$\leqslant \varepsilon_n+\varepsilon_{n-1} \tag{8.140}$$

若 $\varepsilon_n+\varepsilon_{n-1}\leqslant\varepsilon$，这时用 $P_{n(n-1)}(x)$ 作为 $f(x)$ 新的逼近多项式，其次数已降低一次。再由 $P_{n(n-1)}(x)$ 出发，将它表为 $n-1$ 次多项式的形式后，用同样的方法，若 $\varepsilon_n+\varepsilon_{n-1}+\varepsilon_{n-2}\leqslant\varepsilon$，则 $P_{n(n-1)}(x)$ 可再降低方次一次，继续推导，一直做到次数不能再降低为止，此时得到的多项式设为 $P_{nm}(x)$ $(m<n)$，其对于 $f(x)$ 的误差满足

$$\| f(x)-P_{nm}(x)\|_\infty \leqslant \varepsilon_n+\varepsilon_{n-1}+\varepsilon_{n-2}+\cdots+\varepsilon_{n-m}\leqslant\varepsilon \tag{8.141}$$

**例 8.11**　求 $f(x)=e^x$ 在 $[-1,+1]$ 上的三次缩短多项式。

**解**　若取 4 次台劳多项式

$$P_4(x)=1+x+\frac{x^2}{2}+\frac{x^3}{6}+\frac{x^4}{24} \tag{8.142}$$

按式(8.138)得

$$P_{43}(x)=P_4(x)-\frac{1/24}{2^3}T_4(x)$$

$$=0.994\,792+x+0.541\,667x^2+0.166\,667x^3$$

若取 5 次台劳多项式时，可得

$$P_5(x)=1+x+\frac{x^2}{2}+\frac{x^3}{6}+\frac{x^4}{24}+\frac{x^5}{120} \tag{8.143}$$

$$P_{54}(x)=P_5(x)-\frac{1/120}{2^4}T_5(x)$$

$$=1+\frac{383}{384}x+\frac{1}{2}x^2+\frac{17}{96}x^3+\frac{1}{24}x^4$$

在 $P_{54}(x)$ 基础上再降一次

$$P_{53}(x)=P_{54}(x)-\frac{1/24}{2^3}T_4(x)$$

$$=0.994\,792+0.997\,396x+0.541\,667x^2+0.177\,083x^3$$

直接计算

$$\| e^x-P_{43}(x)\|_\infty=0.015\,2,\ \| e^x-P_{53}(x)\|_\infty=0.007\,3$$

由此例可见，$P_{nm}(x)$ 对同一 $m$，次数 $n$ 越大，$P_{nm}(x)$ 逼近 $f(x)$ 的程度越好。

### 5.4　里米兹算法

设 $f(x)\in C[a,b]$，为求其最佳一致逼近多项式 $P_n^*(x)$，由切比雪夫定理看出，关键在于找到具有 $(n+2)$ 个点的切比雪夫交错点组。如果这个点组找到了，那么由 $\Delta^*(x)=f(x)-P_n^*(x)$ 在上述偏差点上的等幅度振荡特性，就可得到以下等幅度振荡方程组

$$\Delta(x_j)=f(x_j)-(a_0^*+a_1^*x_j+\cdots+a_n^*x_j^n)=(-1)^j\sigma E_n(f)$$

$$(\sigma=\pm1, j=0,1,2,\cdots,n+1) \tag{8.144}$$

它是关于 $E_n(f), a_0^*, a_1^*, \cdots, a_n^*$ 的 $n+2$ 阶线性方程组，它有唯一解，从而求得 $P_n^*(x)$ 和 $E_n(f)$。

但是寻找切比雪夫交错点组并非易事，哪怕对于低次的 $P_n^*(x)$ 都很困难。切比雪夫定理启发人们利用极值条件寻求切比雪夫交错点组，里米兹于 1957 年提出了一种逐次校正偏差点的近似求取 $P_n^*(x)$ 的方法，通常称为里米兹算法。

### 5.4.1 里米兹算法步骤

里米兹算法由以下几个步骤组成。

① 首先确定出 $f(x)$在$[a,b]$上的近似最佳一致逼近多项式 $P_n^{(0)}(x)$，一般可取 $f(x)$在$[a,b]$上的截断切比雪夫级数

$$P_n^{(0)}(x)=\frac{1}{2}C_0T_0(x)+C_1T_1(x)+\cdots+C_nT_n(x) \tag{8.145}$$

作为 $P_n^*(x)$的零次近似多项式，并取其极值点集

$$x_j^{(0)}=\frac{a+b}{2}+\frac{b-a}{2}\cos\frac{j\pi}{n+1} \qquad (i=0,1,2,\cdots,n+1) \tag{8.146}$$

作为切比雪夫交错点组的零次近似。

② 令 $P_n^{(r+1)}(x)=\sum_{k=0}^{n}a_k^{(r+1)}x^k$ 为 $P_n^*(x)$ 的 $r+1$ 次近似多项式，其交错点组近似地取为 $x_j^{(r)}(j=0,1,2,\cdots,n+1)$，建立等幅度振荡方程组

$$\Delta_{r+1}(x_j^{(r)})=f(x_j^{(r)})-\sum_{k=0}^{n}a_k^{(r+1)}[x_j^{(r)}]^k=(-1)^j\sigma E_n^{(r+1)}$$
$$(j=0,1,2,\cdots,n+1) \tag{8.147}$$

其中 $E_n^{(r+1)}$ 为$E_n(f)$的 $r+1$ 次近似值。令 $\mu^{(r+1)}=\sigma E_n^{(r+1)}$，则式(8.147)可以改写为

$$\sum_{k=0}^{n}[x_j^{(r)}]^k a_k^{(r+1)}+(-1)^j\mu^{(r+1)}=f(x_j^{(r)}) \qquad (j=0,1,2,\cdots,n+1) \tag{8.148}$$

这是 $n+2$ 个未知数 $a_k^{(r+1)}(k=0,1,2,\cdots,n)$和 $\mu^{(r+1)}$的线性方程组，写成矩阵形式为

$$\begin{bmatrix}1 & x_0^{(r)} & \cdots & [x_0^{(r)}]^n & 1\\ 1 & x_1^{(r)} & \cdots & [x_1^{(r)}]^n & -1\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 1 & x_{n+1}^{(r)} & \cdots & [x_{n+1}^{(r)}]^n & (-1)^{n+1}\end{bmatrix}\begin{bmatrix}a_0^{(r+1)}\\ a_1^{(r+1)}\\ \cdots\\ a_n^{(r+1)}\\ \mu^{(r+1)}\end{bmatrix}=\begin{bmatrix}f(x_0^{(r)})\\ f(x_1^{(r)})\\ \cdots\\ f(x_{n+1}^{(r)})\end{bmatrix}$$

上式的系数矩阵的行列式类似于范德蒙行列式，可以证明，在 $x_j^{(r)}$ 互异的条件下，方程组有唯一解 $a_k^{(r+1)}(k=0,1,2,\cdots,n)$和 $\mu^{(r+1)}$。

③ 在求得 $P_n^{(r+1)}(x)$的基础上，再将误差函数 $\Delta_{r+1}(x)$的局部极值点由 $x_j^{(r)}$ 修改为 $x_j^{(r+1)}$ $(j=0,1,2,\cdots,n+1)$，修改方法见 5.4.2。

④ 用相邻两次近似值之差$|x_j^{(r+1)}-x_j^{(r)}|$或$|a_k^{(r+1)}-a_k^{(r)}|$或$|\mu^{(r+1)}-\mu^{(r)}|$满足一定精度要求为迭代终止条件，若满足终止条件，停止迭代，取 $P_n^{(r+1)}(x)$为 $P_n^*(x)$；否则 $r$ 增 1 后转②。

### 5.4.2 交错点组的调整方法

关于由 $x_j^{(r)}(j=0,1,2,\cdots,n+1)$调整为 $x_j^{(r+1)}(j=0,1,2,\cdots,n+1)$的方法有以下两种。

(1) 单点交换法

它用达到$\|\Delta_{r+1}(x)\|_\infty$的偏差点 $x'$取代 $x_j^{(r)}(j=0,1,2,\cdots,n+1)$中的某一点而其余点保持不变来构成新交错点组的方法。具体处理法则如下。

① 当 $x'\in[a,x_0^{(r)})$时，若 $\Delta_{r+1}(x_0^{(r)})$与 $\Delta_{r+1}(x')$同号，则用 $x'$代替 $x_0^{(r)}$ 而其余点保持不变而构成新的交错点组$\{x_j^{(r+1)}\}$；若异号，则取新交错点组为$\{x_j^{(r+1)}\}=\{x',x_0^{(r)},x_1^{(r)},\cdots,x_n^{(r)}\}$。

② 当 $x'\in(x_{n+1}^{(r)},b]$时，若 $\Delta_{r+1}(x_{n+1}^{(r)})$与 $\Delta_{r+1}(x')$同号，则用 $x'$代替 $x_{n+1}^{(r)}$而其余点保持不变而构

成新的交错点组$\{x_i^{(r+1)}\}$；若异号，则取新交错点组为$\{x_j^{(r+1)}\}=\{x_1^{(r)},x_2^{(r)},\cdots,x_{n+1}^{(r)},x'\}$。

③ 当$x'\in(x_j^{(r)},x_{j+1}^{(r)})$时，若$\Delta_{r+1}(x_j^{(r)})$与$\Delta_{r+1}(x')$同号，则用$x'$代替$x_j^{(r)}$；否则用$x'$代替$x_{j+1}^{(r)}$而其余点保持不变来构成新的交错点组$\{x_j^{(r+1)}\}$。

(2)　多点交换法

这种方法使用$\Delta_{r+1}(x)$的多个局部极值点或局部最大值点分别按单点交换法中法则依次代替$\{x_j^{(r)}\}$中的某些点或全部点来构成新交错点组的方法。它比单点交换法有更高的效率。

关于$\Delta_{r+1}(x)$在$[a,b]$上的局部极值点或局部最大值点的求取方法可采用以下几种方法。

① 切线法：

设$\Delta_{r+1}(x)$在$[a,b]$上有连续的二阶导数，则极值点应满足

$$\Delta'_{r+1}(x)=0 \tag{8.149}$$

应用切线法于$x_j^{(r)}$点得$x_j^{(r)}$近旁的近似局部极值点

$$x'=x_j^{(r)}-\frac{\Delta'_{r+1}(x_j^{(r)})}{\Delta''_{r+1}(x_j^{(r)})}=x_j^{(r)}+\delta x_j^{(r)} \tag{8.150}$$

上述公式只适用于内偏差点的场合。

② 抛物线法：

对于每一个$x_j^{(r)}$点来说，若其为内点时，取

$$\begin{cases}u=x_j^{(r)}\\ v=x_j^{(r)}+\alpha(x_{j+1}^{(r)}-x_j^{(r)})\\ w=x_j^{(r)}-\alpha(x_j^{(r)}-x_{j-1}^{(r)})\end{cases} \tag{8.151}$$

其中$0<\alpha<1$。对于$x_0^{(r)}$，它可能为内点亦可能为左端点$a$，可如下取法

$$\begin{cases}u=x_0^{(r)}\\ v=x_0^{(r)}+\alpha(x_1^{(r)}-x_0^{(r)})\\ w=\begin{cases}x_0^{(r)}+\beta(x_1^{(r)}-x_0^{(r)}), & x_0^{(r)}=a\\ x_0^{(r)}-\alpha(x_0^{(r)}-a), & x_0^{(r)}>a\end{cases}\end{cases} \tag{8.152}$$

其中$0<\alpha,\beta<1,\alpha\neq\beta$；同法，对$x_{n+1}^{(r)}$来说可取

$$\begin{cases}u=x_{n+1}^{(r)}\\ v=x_{n+1}^{(r)}-\alpha(x_{n+1}^{(r)}-x_n^{(r)})\\ w=\begin{cases}x_{n+1}^{(r)}-\beta(x_{n+1}^{(r)}-x_n^{(r)}), & x_{n+1}^{(r)}=b\\ x_{n+1}^{(r)}+\alpha(b-x_{n+1}^{(r)}), & x_{n+1}^{(r)}<b\end{cases}\end{cases} \tag{8.153}$$

建立如下数值表表8.9。

**表 8.9**

| $x$ | $u$ | $v$ | $w$ |
|---|---|---|---|
| $\Delta_{r+1}(x)$ | $\Delta_{r+1}(u)$ | $\Delta_{r+1}(v)$ | $\Delta_{r+1}(w)$ |

和拉格朗日插值公式

$$L_2(x)=\frac{(x-u)(x-v)}{(w-u)(w-v)}\Delta_{r+1}(w)+\frac{(x-v)(x-w)}{(u-v)(u-w)}\Delta_{r+1}(u)+\frac{(x-u)(x-w)}{(v-u)(v-w)}\Delta_{r+1}(v) \tag{8.154}$$

为求极值点，对 $L_2(x)$ 求导得

$$L'_2(x)=\frac{2x-u-v}{(w-u)(w-v)}\Delta_{r+1}(w)+\frac{2x-v-w}{(u-v)(u-w)}\Delta_{r+1}(u)+\frac{2x-u-w}{(v-u)(v-w)}\Delta_{r+1}(v)$$

对上式用 $(w-u)(u-v)(v-w)$ 乘后令其为 0 得

$$(2x-u-v)(u-v)\Delta_{r+1}(w)+(2x-v-w)(v-w)\Delta_{r+1}(u)+(2x-u-w)(w-u)\Delta_{r+1}(v)=0$$

即

$$2x[(u-v)\Delta_{r+1}(w)+(v-w)\Delta_{r+1}(u)+(w-u)\Delta_{r+1}(v)]=(u^2-v^2)\Delta_{r+1}(w)+(v^2-w^2)\Delta_{r+1}(u)+(w^2-u^2)\Delta_{r+1}(v)$$

解得极值点为

$$x_j^{(r+1)}=\frac{1}{2}\frac{(u^2-v^2)\Delta_{r+1}(w)+(v^2-w^2)\Delta_{r+1}(u)+(w^2-u^2)\Delta_{r+1}(v)}{(u-v)\Delta_{r+1}(w)+(v-w)\Delta_{r+1}(u)+(w-u)\Delta_{r+1}(v)}=\frac{1}{2}(u+v)+\frac{(u-w)(w-v)}{2d}[\Delta_{r+1}(u)-\Delta_{r+1}(v)] \tag{8.155}$$

其中 $d=(u-v)\Delta_{r+1}(w)+(v-w)\Delta_{r+1}(u)+(w-u)\Delta_{r+1}(v)$

当 $u,v,w$ 的值十分接近时，$d$ 的值很小或等于零，这时运算误差较大，可处理如下

$$x_j^{(r+1)}=\begin{cases}\frac{1}{2}(u+v),d=0\\ \frac{1}{2}(u+v)+\frac{(u-w)(w-v)}{2d}[\Delta_{r+1}(u)-\Delta_{r+1}(v)],d\neq 0\ (0\leqslant j\leqslant n+1)\end{cases} \tag{8.156}$$

当上述 $x_j^{(r+1)}$ 算得后，还要检查 $x_j^{(r+1)}\in[a,b]$ 否？当 $x_j^{(r+1)}\overline{\in}[a,b]$ 时，$x'$ 应由 $u,v,w$ 三个中按其对应的 $|\Delta_{r+1}(x)|$ 为最大的原则来选定；当 $x_j^{(r+1)}\in[a,b]$ 时，$x'$ 应由 $u,v,w,x_j^{(r+1)}$ 四个中按其对应的 $|\Delta_{r+1}(x)|$ 为最大的原则选定。在极小值或局部最小值的情况下，可仿此处理。

③ 成功-失败法：

此法根据 $x_j^{(r)}$ 为极大或极小值点特性确定成功的标志。在极小值或局部最小值的情况下，成功的判别标志为下山条件，否则为上山条件。以极小值为例，为求 $\Delta_{r+1}(x)$ 的极小值或局部最小值，设从 $x_j^{(r)}=A$ 出发，适当选择步长 $h$，计算 $\Delta_{r+1}(A)$ 与 $\Delta_{r+1}(A+h)$ 并处理如下：当

$$\Delta_{r+1}(A+h)<\Delta_{r+1}(A) \qquad (\text{下山条件}) \tag{8.157}$$

满足时，将 $A+h$ 作为新的 $A$，步长 $h$ 改为 $\alpha h(\alpha>1)$；当下山条件不满足时，$A$ 不变，步长改为 $-\beta h(0<\beta<1)$。继续计算 $\Delta_{r+1}(A)$ 与 $\Delta_{r+1}(A+h)$，同法检验下山条件及进行有关处理，直到 $h$ 的值在某个给定的小范围内为止，此时可认为 $x'=A$ 是极小值点或局部最小值点。

### 5.4.3 里米兹算法的收敛性

设 $P_n^*(x)$ 的零次近似多项式为 $P_n^{(0)}(x)$，其误差函数在零次交错点组 $x_j^{(0)}(j=0,1,2,\cdots,n+1)$ 上的值为

$$\Delta_0(x_j^{(0)})=f(x_j^{(0)})-P_n^{(0)}(x_j^{(0)}) \qquad (j=0,1,2,\cdots,n+1) \tag{8.158}$$

另设 $P_n^{(1)}(x)$ 为 $P_n^*(x)$ 的一次近似多项式，其误差函数 $\Delta_1(x)$ 在零次交错点组 $\{x_j^{(0)}\}$ 上的等幅度振荡方程组为

$$\Delta_1(x_j^{(0)}) = f(x_j^{(0)}) - P_n^{(1)}(x_j^{(0)}) = (-1)^j \sigma E_n^{(1)} \tag{8.159}$$

由式(8.158)知

$$f(x_j^{(0)}) = \Delta_0(x_j^{(0)}) + P_n^{(0)}(x_j^{(0)})$$

代入式(8.159)后整理得

$$\Delta_0(x_j^{(0)}) - [P_n^{(1)}(x_j^{(0)}) - P_n^{(0)}(x_j^{(0)})] = \Delta_0(x_j^{(0)}) - \sum_{k=0}^{n}(x_j^{(0)})^k \Delta a_k^{(0)} = (-1)^j \sigma E_n^{(1)}$$

或

$$(-1)^j \sigma E_n^{(1)} + \sum_{k=0}^{n}(x_j^{(0)})^k \Delta a_k^{(0)} = \Delta_0(x_j^{(0)}) \quad (j=0,1,2,\cdots,n+1) \tag{8.160}$$

上式可用矩阵形式表为

$$\begin{bmatrix} 1 & 1 & x_0^{(0)} & (x_0^{(0)})^2 & \cdots & (x_0^{(0)})^n \\ -1 & 1 & x_1^{(0)} & (x_1^{(0)})^2 & \cdots & (x_1^{(0)})^n \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ (-1)^{n+1} & 1 & x_{n+1}^{(0)} & (x_{n+1}^{(0)})^2 & \cdots & (x_{n+1}^{(0)})^n \end{bmatrix} \begin{bmatrix} \sigma E_n^{(1)} \\ \Delta a_0^{(0)} \\ \cdots \\ \Delta a_n^{(0)} \end{bmatrix} = \begin{bmatrix} \Delta_0(x_0^{(0)}) \\ \Delta_0(x_1^{(0)}) \\ \cdots \\ \Delta_0(x_{n+1}^{(0)}) \end{bmatrix}$$

解出 $\sigma E_n^{(1)}$ 得以下表达式

$$\sigma E_n^{(1)} = \frac{\begin{vmatrix} \Delta_0(x_0^{(0)}) & 1 & x_0^{(0)} & (x_0^{(0)})^2 & \cdots & (x_0^{(0)})^n \\ \Delta_0(x_1^{(0)}) & 1 & x_1^{(0)} & (x_1^{(0)})^2 & \cdots & (x_1^{(0)})^n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Delta_0(x_{n+1}^{(0)}) & 1 & x_{n+1}^{(0)} & (x_{n+1}^{(0)})^2 & \cdots & (x_{n+1}^{(0)})^n \end{vmatrix}}{\begin{vmatrix} 1 & 1 & x_0^{(0)} & (x_0^{(0)})^2 & \cdots & (x_0^{(0)})^n \\ -1 & 1 & x_1^{(0)} & (x_1^{(0)})^2 & \cdots & (x_1^{(0)})^n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (-1)^{n+1} & 1 & x_{n+1}^{(0)} & (x_{n+1}^{(0)})^2 & \cdots & (x_{n+1}^{(0)})^n \end{vmatrix}}$$

上式按第一列元素展开行列式得

$$\sigma E_n^{(1)} = \frac{D_0\Delta_0(x_0^{(0)}) - D_1\Delta_0(x_1^{(0)}) + D_2\Delta_0(x_2^{(0)}) + \cdots + (-1)^{n+1}D_{n+1}\Delta_0(x_{n+1}^{(0)})}{D_0 + D_1 + \cdots + D_{n+1}} \tag{8.161}$$

式中，$D_i$ 是范德蒙行列式，这种行列式可以分解为不为零的因子$(x_q^{(0)} - x_p^{(0)})$之积。例如

$$D_0 = \begin{vmatrix} 1 & x_1^{(0)} & (x_1^{(0)})^2 & \cdots & (x_1^{(0)})^n \\ 1 & x_2^{(0)} & (x_2^{(0)})^2 & \cdots & (x_2^{(0)})^n \\ & & \cdots & & \\ 1 & x_{n+1}^{(0)} & (x_{n+1}^{(0)})^2 & \cdots & (x_{n+1}^{(0)})^n \end{vmatrix}$$

可以分解成下列诸因子之积

$$\begin{aligned} &(x_{n+1}^{(0)} - x_n^{(0)})(x_{n+1}^{(0)} - x_{n-1}^{(0)})\cdots(x_{n+1}^{(0)} - x_2^{(0)})(x_{n+1}^{(0)} - x_1^{(0)}) \\ &\qquad (x_n^{(0)} - x_{n-1}^{(0)})\cdots(x_n^{(0)} - x_2^{(0)})(x_n^{(0)} - x_1^{(0)}) \\ &\qquad\qquad \cdots \\ &\qquad\qquad\qquad (x_3^{(0)} - x_2^{(0)})(x_3^{(0)} - x_1^{(0)}) \\ &\qquad\qquad\qquad\qquad (x_2^{(0)} - x_1^{(0)}) \end{aligned}$$

即 $D_0 = \prod_{1 \leqslant p < q \leqslant n+1}(x_q^{(0)} - x_p^{(0)}) > 0$。仿之可求得

$$D_i=\prod_{0\leqslant p<q\leqslant n+1}{}'(x_q^{(0)}-x_p^{(0)})>0,\quad i=1,2,\cdots,n+1$$

其中$\prod'$表示式中缺$x_i^{(0)}$项(正如$D_0$中缺$x_0^{(0)}$项一样)。今对(8.161)式的分子分母同用$\prod\limits_{0\leqslant p<q\leqslant n+1}(x_q^{(0)}-x_p^{(0)})$除之,得

$$\sigma E_n^{(1)}=\frac{d_0\Delta_0(x_0^{(0)})-d_1\Delta_0(x_1^{(0)})+d_2\Delta_0(x_2^{(0)})+\cdots+(-1)^{n+1}d_{n+1}\Delta_0(x_{n+1}^{(0)})}{d_0+d_1+\cdots+d_{n+1}}\tag{8.162}$$

式中

$$d_j=\frac{1}{(x_j^{(0)}-x_0^{(0)})(x_j^{(0)}-x_1^{(0)})\cdots(x_j^{(0)}-x_{j-1}^{(0)})(x_{j+1}^{(0)}-x_j^{(0)})\cdots(x_{n+1}^{(0)}-x_j^{(0)})}\tag{8.163}$$

显见$d_j>0$。由于$\Delta_0(x)$在交错点组$\{x_j^{(0)}\}$上的符号交替改变,所以(8.162)式可化为

$$E_n^{(1)}=\frac{\sum\limits_{j=0}^{n+1}d_j|\Delta_0(x_j^{(0)})|}{\sum\limits_{j=0}^{n+1}d_j},\qquad\sigma=\operatorname{sgn}\Delta_0(x_0^{(0)})\tag{8.164}$$

上式表明,$E_n^{(1)}$是对$|\Delta_0(x_0^{(0)})|,|\Delta_0(x_1^{(0)})|,\cdots,|\Delta_0(x_{n+1}^{(0)})|$加权重$d_0,d_1,\cdots,d_{n+1}$后的算术平均值,因此这个平均值必满足

$$A_0<E_n^{(1)}<L_0\tag{8.165}$$

其中 $$A_0=\min_{0\leqslant j\leqslant n+1}|\Delta_0(x_j^{(0)})|,\quad L_0=\max_{0\leqslant j\leqslant n+1}|\Delta_0(x_j^{(0)})|$$

同样 $$A_1=\min_{0\leqslant j\leqslant n+1}|\Delta_1(x_j^{(1)})|,\quad L_1=\max_{0\leqslant j\leqslant n+1}|\Delta_1(x_j^{(1)})|$$

因$E_n^{(1)}$是$\Delta_1(x)$在零次交错点组$x_j^{(0)}(j=0,1,2,\cdots n+1)$上的函数绝对值,一般$x_j^{(0)}(j=0,1,2,\cdots,n+1)$不是$\Delta_1(x)$的极值点,所以应有以下不等式

$$E_n^{(1)}\leqslant A_1\tag{8.166}$$

综合式(8.165)和式(8.166)得

$$A_0<E_n^{(1)}\leqslant A_1\tag{8.167}$$

一般应有

$$A_i<E_n^{(i+1)}<L_i\tag{8.168}$$

$$A_i<E_n^{(i+1)}<A_{i+1}\tag{8.169}$$

其中 $$A_i=\min_{0\leqslant j\leqslant n+1}|\Delta_i(x_j^{(i)})|,L_i=\max_{0\leqslant j\leqslant n+1}|\Delta_i(x_j^{(i)})|$$

由于$E_n^{(i+1)}$是$P_n^{(i+1)}(x)$在$i$次交错点组$x_j^{(i)}(j=0,1,2,\cdots,n+1)$上的最佳逼近值,记$i$次交错点组为$s$,则有以下关系式

$$\begin{aligned}E_n^{(i+1)}&=\max_{x\in s}|f(x)-P_n^{(i+1)}(x)|\\&\leqslant\max_{x\in s}|f(x)-P_n^*(x)|\text{(因为}P_n^{(i+1)}(x)\text{是}s\text{上的最佳一致逼近多项式)}\\&\leqslant\max_{a\leqslant x\leqslant b}|f(x)-P_n^*(x)|\text{(局部极大值<整体极大值)}\\&=E_n(f)\end{aligned}$$

由此获得以下不等式

$$E_n^{(i+1)}\leqslant E_n(f)\tag{8.170}$$

当$s$为切比雪夫交错点组时,上式中的等号成立。因$\Delta_i(x)$并非$\Delta^*(x)$,显然有以下不等式

$$E_n(f)<L_i\qquad(i=0,1,2,\cdots)\tag{8.171}$$

成立。

在以上关系式的基础上，我们来证明里米兹算法的收敛性。由关系式(8.168)知，必存在有 $0<\theta<1$，使满足下式

$$E_n^{(i+1)}-A_i>(1-\theta)(L_i-A_i) \tag{8.172}$$

利用式(8.169)和式(8.171)可将上式化为

$$A_{i+1}-A_i>(1-\theta)(E_n(f)-A_i)$$

即得

$$E_n(f)-A_{i+1}<\theta[E_n(f)-A_i]$$

反复利用上式可得

$$E_n(f)-A_{i+1}<\theta[E_n(f)-A_i]<\theta^2[E_n(f)-A_{i-1}]<\cdots<\theta^{i+1}[E_n(f)-A_0]$$

对上式取极限可得

$$\lim_{i\to\infty}A_{i+1}=E_n(f) \tag{8.173}$$

由式(8.168)、式(8.170)、式(8.172)可得以下不等式

$$L_i-E_n(f)<L_i-A_i<\frac{E_n^{(i+1)}-A_i}{1-\theta}<\frac{E_n(f)-A_i}{1-\theta}$$

推知

$$\lim_{i\to\infty}L_i=E_n(f) \tag{8.174}$$

综合式(8.173)与式(8.174)知，用里米兹算法获得的近似多项式 $P_n^{(0)}(x)$，$P_n^{(1)}(x)\cdots$收敛于最佳一致逼近多项式 $P_n^*(x)$。

**例 8.12**　求 $f(x)=\sqrt{x}$在$[\frac{1}{4},1]$上的一次最佳一致逼近多项式 $P_1^*(x)=a_0^*+a_1^*x$。

**解**　令

$$x=\frac{1-\frac{1}{4}}{2}t+\frac{1+\frac{1}{4}}{2}=\frac{3}{8}t+\frac{5}{8}=0.375t+0.625 \tag{8.175}$$

其中$-1\leqslant t\leqslant+1$。则

$$f(x)=f(0.375t+0.625)=F(t)=\sqrt{0.375t+0.625} \tag{8.176}$$

设 $F(t)$在$[-1,+1]$上的一次最佳一致逼近多项式为 $P_1^*(t)=b_0^*+b_1^*t$，它的零次近似多项式可取下列 $F(t)$的切比雪夫多项式展开式

$$P_1^{(0)}(t)=\frac{b_0^{(0)}}{2}+b_1^{(0)}T_1(t) \tag{8.177}$$

取 $T_2(t)$的零点

$$t_i=\cos\frac{2i-1}{4}\pi\qquad(i=1,2) \tag{8.178}$$

计算得

$$t_1=\cos\frac{\pi}{4}=0.707\ 11,\ t_2=\cos\frac{3}{4}\pi=-0.707\ 11$$

由式(8.95)中取 $n=1$ 算得

$$b_0^{(0)}=\sqrt{0.375\times0.707\ 11+0.625}+\sqrt{0.375\times(-0.707\ 11)+0.625}=1.543\ 34$$

$$b_1^{(0)}=0.707\ 11\times\sqrt{0.375\times0.707\ 11+0.625}+(-0.707\ 11)\times\sqrt{0.375\times(-0.707\ 11)+0.625}=0.242\ 98$$

所以

$$P_1^{(0)}(t)=0.771\ 67+0.242\ 98t \tag{8.179}$$

$$\Delta_0(t)=\sqrt{0.375t+0.625}-0.771\ 67-0.242\ 98t \tag{8.180}$$

今取 $T_2(t)$的极值点

$$t_j = \cos \frac{j}{2}\pi \qquad (j = 0,1,2) \tag{8.181}$$

作为 $\Delta_0(t)$的交错点组,计算得

$$t_0^{(0)} = \cos 0 = 1, \quad t_1^{(0)} = \cos \frac{\pi}{2} = 0, \quad t_2^{(0)} = \cos \pi = -1 \tag{8.182}$$

因 $$F''(t) = -0.035\ 156\ 25(0.375t+0.625)^{-1.5}$$

在区间[−1,+1]上不变号,所以误差函数是标准的误差函数,则端点±1 均是偏差点且偏差点个数为 1+2=3 个。令 $P_1^*(t)$的一次近似多项式为 $P_1^{(1)}(t)=b_0^{(1)}+b_1^{(1)}t$,在交错点组(8.182) 上建立 $\Delta_1(x)$的等幅度振荡方程组

$$\begin{cases} \sqrt{0.375\times(-1)+0.625}-[b_0^{(1)}+b_1^{(1)}\times(-1)]=\mu_1^{(1)} \\ \sqrt{0.375\times 0+0.625}-(b_0^{(1)}+b_1^{(1)}\times 0)=-\mu_1^{(1)} \\ \sqrt{0.375\times 1+0.625}-(b_0^{(1)}+b_1^{(1)}\times 1)=\mu_1^{(1)} \end{cases} \tag{8.183}$$

式中,$\mu_1^{(1)}$ 为 $E_1^{(1)}$ 的代数值,它等于$+E_1^{(1)}$ 或$-E_1^{(1)}$。求解(8.183)式得

$$b_0^{(1)} = 0.770\ 28, \quad b_1^{(1)} = 0.25, \quad \mu_1^{(1)} = -0.020\ 28$$

$$P_1^{(1)}(t) = 0.770\ 28 + 0.25t$$

按式(8.150)得

$$t_1^{(r+1)} = t_1^{(r)} + \delta t_1^{(r)} \qquad (r = 0,1,2,\cdots) \tag{8.184}$$

其中 $$\delta t_1^{(r)} = -\frac{\dfrac{0.187\ 5}{\sqrt{0.375t_1^{(r)}+0.625}}-0.25}{-0.035\ 156\ 25\times(0.375t_1^{(r)}+0.625)^{-1.5}}$$

$$= -\frac{(0.375t_1^{(r)}+0.625)(0.25\sqrt{0.375t_1^{(r)}+0.625}-0.187\ 5)}{0.035\ 156\ 25} \tag{8.185}$$

以 $t_1^{(0)}$ 代入上式得 $\delta t_1^{(0)}=-0.180\ 31$,所以

$$\begin{cases} t_1^{(1)} = t_1^{(0)} + \delta t_1^{(0)} = 0 - 0.180\ 31 = -0.180\ 31 \\ t_0^{(1)} = -1 \\ t_2^{(1)} = +1 \end{cases} \tag{8.186}$$

令 $P_1^{(2)}(t)=b_0^{(2)}+b_1^{(2)}t$,建立 $\Delta_2(x)$在交错点组(8.186)上的等幅度振荡方程组

$$\begin{cases} \sqrt{0.375\times(-1)+0.625}-[b_0^{(2)}+b_1^{(2)}\times(-1)]=\mu_1^{(2)} \\ \sqrt{0.375\times(-0.180\ 31)+0.625}-[b_0^{(2)}+b_1^{(2)}\times(-0.180\ 31)]=-\mu_1^{(2)} \\ \sqrt{0.375\times 1+0.625}-(b_0^{(2)}+b_1^{(2)}\times 1)=\mu_1^{(2)} \end{cases} \tag{8.187}$$

解得 $b_0^{(2)}=0.770\ 83$, $b_1^{(2)}=0.25$, $\mu_1^{(2)}=-0.020\ 83$。则

$$P_1^{(2)}(t) = 0.770\ 83 + 0.25t$$

计算 $$\delta t_1^{(1)}=0.013\ 48$$

$$\begin{cases} t_1^{(2)} = t_1^{(1)} + \delta t_1^{(1)} = -0.180\ 31+0.013\ 48 = -0.166\ 83 \\ t_0^{(2)} = -1 \\ t_2^{(2)} = 1 \end{cases} \tag{8.188}$$

令 $P_1^{(3)}(x)=b_0^{(3)}+b_1^{(3)}t$,建立 $\Delta_3(x)$在交错点组(8.188)上的等幅度振荡方程组,同法可解

得 $b_0^{(3)}=0.770\ 83, b_1^{(3)}=0.25, \mu_1^{(3)}=-0.020\ 83$。这时系数的数值已稳定到小数后第五位，取

$$P_1^*(x)=0.770\ 83+0.25t \tag{8.189}$$

以

$$t=\frac{x-0.625}{0.375}$$

代入(8.189)式得

$$P_1^*(x)=0.770\ 83+0.25\times\frac{x-0.625}{0.375}=0.666\ 67+0.354\ 16x$$

# 习题八

8.1 试用最小平方逼近法，求解下列超定方程组

$$\begin{cases} x_1 + 2x_2 = 4 \\ 2x_1 + x_2 = 5 \\ 2x_1 + 2x_2 = 6 \\ -x_1 + 2x_2 = 2 \\ 3x_1 - x_2 = 4 \end{cases}$$

8.2 求下列函数在指定区间上的一次最小平方逼近多项式。

(1) $f(x)=x^2-2x+3$, $[0,1]$

(2) $f(x)=\frac{1}{x}$, $[1,3]$

(3) $f(x)=e^{-x}$, $[0,1]$

(4) $f(x)=\cos\pi x$, $[0,1]$

8.3 给定数据表

| $x$ | 0 | 0.15 | 0.31 | 0.5 | 0.6 | 0.75 |
|---|---|---|---|---|---|---|
| $y$ | 1.0 | 1.004 | 1.031 | 1.117 | 1.223 | 1.422 |

求1,2,3,4次最小平方逼近多项式。

8.4 给定数据表

| $x$ | 0.000 | 1.445 | 2.890 | 4.335 | 5.780 |
|---|---|---|---|---|---|
| $y$ | 1.841 9 | 2.963 3 | 18.236 0 | 98.741 0 | 529.217 8 |

求形如 $ae^{bx}$ 的最小平方拟合函数。

8.5 用正交多项式的最小平方逼近方法，求下列函数的三次多项式拟合曲线。

(1) $f(x)=\cos x$，　$x_j=0.2j$　$(j=0,1,2,3,4,5)$

(2) $f(x)=\ln x$，　$x_j=1+0.2j$　$(j=0,1,2,3,4,5)$

(3) $f(x)=e^{-0.1x}$，　$x_j=0.2j$　$(j=0,1,2,3,4,5)$

8.6 设在 $0\leqslant x\leqslant 1$ 上给定 $P(x)=1-x+x^2-x^3+x^4$，试在容许误差0.008的要求下降低 $P(x)$ 的次数。

8.7 求函数 $f(x)=\arcsin x$ 在 $x\in[-1,+1]$ 上按切比雪夫多项式的展开式。

8.8 利用切比雪夫多项式零点插值法，求下列函数的三次近似最佳一致逼近多项式。

(1) $f(x)=\arctan x$，　$x\in[-1,+1]$

(2) $f(x)=xe^{-x}$，　$x\in[0,1.5]$

8.9 用缩短幂级数法，求 $f(x)=\sin x$ 在 $[-1,1]$ 上的三次多项式，使误差 $\leqslant 0.005$。

8.10 求函数 $f(x)=\sqrt{1+x^2}$ 在 $[0,1]$ 上的一次最佳一致逼近多项式。

8.11 设在区间$[-1,+1]$上 $f(x)=1-\frac{1}{2}x-\frac{1}{8}x^2-\frac{3}{24}x^3-\frac{15}{385}x^4-\frac{165}{3\,840}x^5$，试将 $f(x)$ 降低到三次多项式，并估计误差。

8.12 设 $f(x)=x^4+3x^3-1$，在$[0,1]$上求三次最佳一致逼近多项式。

8.13 求 $\sin x$ 在$[0,\frac{\pi}{2}]$上的一次最佳一致逼近多项式，并估计误差。

8.14 求 $e^x$ 在$[0,1]$上的一次最佳一致逼近多项式，并估计误差。

# 第九章　矩阵特征值和特征向量的计算

$n$ 阶方阵 $\boldsymbol{A}$ 的特征值问题由

$$\boldsymbol{AX}=\lambda\boldsymbol{X} \tag{9.1}$$

或

$$[\boldsymbol{A}-\lambda\boldsymbol{I}]\boldsymbol{X}=0 \tag{9.2}$$

确定，式中 $\boldsymbol{A}-\lambda\boldsymbol{I}$ 称为 $\boldsymbol{A}$ 的特征矩阵，$\boldsymbol{I}$ 为单位矩阵。满足(9.1)或(9.2)的 $\lambda$ 和非零向量 $\boldsymbol{X}$ 分别称为矩阵 $\boldsymbol{A}$ 的特征值和相应于 $\lambda$ 的特征向量。因(9.2)是一个齐次线性方程组，它具有非零解的充要条件是下列特征行列式等于零

$$\det[\boldsymbol{A}-\lambda\boldsymbol{I}]=\begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ \vdots & & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{vmatrix}=0 \tag{9.3}$$

将上式的行列式展开后可得到以下代数方程

$$\begin{aligned} f(\lambda) &= (-1)^n[\lambda^n-\sigma_1\lambda^{n-1}+\sigma_2\lambda^{n-2}+\cdots+(-1)^n\sigma_n] \\ &= a_0\lambda^n+a_1\lambda^{n-1}+\cdots+a_{n-1}\lambda+a_n=0 \end{aligned} \tag{9.4}$$

称为特征方程，其中

$$\begin{cases} \sigma_1=\sum\limits_{i=1}^{n}a_{ii} \\ \sigma_2=\sum\limits_{i<j}\begin{vmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{vmatrix}, \quad \sigma_3=\sum\limits_{i<j<k}\begin{vmatrix} a_{ii} & a_{ij} & a_{ik} \\ a_{ji} & a_{jj} & a_{jk} \\ a_{ki} & a_{kj} & a_{kk} \end{vmatrix} \\ \cdots \\ \sigma_n=|\boldsymbol{A}| \end{cases} \tag{9.5}$$

它们分别是对角线元素的所有一阶主子式之和；所有二阶主子式之和直到 $n$ 阶主子式的和。

表面上看，只要求出方程(9.4)的根，进而求取方程(9.2)的解 $\boldsymbol{X}$ 就行了。这对于 $n$ 是很小的整数是可行的。只要 $n$ 稍大，计算工作的困难程度就会迅速增加。并且由于计算误差，展开式(9.4)未必就是精确的特征方程，自然更不必说求解式(9.4)与式(9.2)的繁重了。

本章仅对实矩阵进行讨论。

## §1　幂法和反幂法

### 1.1　幂法

幂法主要用来求一个矩阵的模为最大的特征值和相应的特征向量的方法，它虽不是一个一般的方法，但在许多场合是有用的。

假定 $|\boldsymbol{A}|$ 有 $n$ 个线性独立的特征向量 $\boldsymbol{V}_j(j=1,2\cdots,n)$，它们形成 $n$ 维空间的基，则该空间

的任何向量 $\boldsymbol{X}$ 可唯一地表为

$$\boldsymbol{X} = c_1\boldsymbol{V}_1 + c_2\boldsymbol{V} + \cdots + c_n\boldsymbol{V}_n \tag{9.6}$$

取初始向量为 $\boldsymbol{X}_0$，并也有表达式

$$\boldsymbol{X}_0 = \alpha_1\boldsymbol{V}_1 + \alpha_2\boldsymbol{V}_2 + \cdots + \alpha_n\boldsymbol{V}_n \tag{9.7}$$

计算

$$\boldsymbol{X}_{k+1} = \boldsymbol{A}\boldsymbol{X}_k \qquad (k = 0,1,2\cdots) \tag{9.8}$$

获得迭代序列 $\{\boldsymbol{X}_k\}$，然后，依据迭代序列中存在于迭代值间的关系，求取按模最大特征值及特征向量的近似值。以下分不同的情况进行分析。

(1) $|\lambda_1| > |\lambda_2| \geqslant |\lambda_3| \geqslant \cdots \geqslant |\lambda_n|$ 情况

任取一初始向量

$$\boldsymbol{X}_0 = \alpha_1\boldsymbol{V}_1 + \alpha_2\boldsymbol{V}_2 + \cdots + \alpha_n\boldsymbol{V}_n$$

计算迭代序列

$$\boldsymbol{X}_{k+1} = \boldsymbol{A}\boldsymbol{X}_k \qquad (k = 0,1,2,\cdots)$$

因为

$$\begin{aligned}
\boldsymbol{X}_1 &= \boldsymbol{A}\boldsymbol{X}_0 = \alpha_1\boldsymbol{A}\boldsymbol{V}_1 + \alpha_2\boldsymbol{A}\boldsymbol{V}_2 + \cdots + \alpha_n\boldsymbol{A}\boldsymbol{V}_n \\
&= \alpha_1\lambda_1\boldsymbol{V}_1 + \alpha_2\lambda_2\boldsymbol{V}_2 + \cdots + \alpha_n\lambda_n\boldsymbol{V}_n \\
&= \lambda_1\left[\alpha_1\boldsymbol{V}_1 + \alpha_2\frac{\lambda_2}{\lambda_1}\boldsymbol{V}_2 + \cdots + \alpha_n\frac{\lambda_n}{\lambda_1}\boldsymbol{V}_n\right] \\
\boldsymbol{X}_2 &= \boldsymbol{A}\boldsymbol{X}_1 = \lambda_1\left[\alpha_1\boldsymbol{A}\boldsymbol{V}_1 + \alpha_2\frac{\lambda_2}{\lambda_1}\boldsymbol{A}\boldsymbol{V}_2 + \cdots + \alpha_n\frac{\lambda_n}{\lambda_1}\boldsymbol{A}\boldsymbol{V}_n\right] \\
&= \lambda_1\left[\alpha_1\lambda_1\boldsymbol{V}_1 + \alpha_2\frac{\lambda_2}{\lambda_1}\lambda_2\boldsymbol{V}_2 + \cdots + \alpha_n\frac{\lambda_n}{\lambda_1}\lambda_n\boldsymbol{V}_n\right] \\
&= \lambda_1^2\left[\alpha_1\boldsymbol{V}_1 + \alpha_2\left(\frac{\lambda_2}{\lambda_1}\right)^2\boldsymbol{V} + \cdots + \alpha_n\left(\frac{\lambda_n}{\lambda_1}\right)^2\boldsymbol{V}_n\right] \\
&\cdots \\
\boldsymbol{X}_k &= \lambda_1^k\left[\alpha_1\boldsymbol{V}_1 + \alpha_2\left(\frac{\lambda_2}{\lambda_1}\right)^k\boldsymbol{V}_2 + \cdots + \alpha_n\left(\frac{\lambda_n}{\lambda_1}\right)^k\boldsymbol{V}_n\right] \\
&\approx \lambda_1^k\alpha_1\boldsymbol{V}_1 \quad (k\text{ 足够大时})
\end{aligned}$$

同样

$$\boldsymbol{X}_{k+1} \approx \lambda_1^{k+1}\alpha_1\boldsymbol{V}_1$$

则 $\dfrac{\boldsymbol{X}_{k+1}}{\boldsymbol{X}_k} \xrightarrow{k\to\infty} \lambda_1$（即两个向量任何一个分量比的极限都为 $\lambda_1$）　(9.9)

计算中的有关问题如下。

① 迭代的终止条件。

对于 $\boldsymbol{X}_k = [(\boldsymbol{X}_k)_1, (\boldsymbol{X}_k)_2, \cdots, (\boldsymbol{X}_k)_n]$ 和 $\boldsymbol{X}_{k+1} = [(\boldsymbol{X}_{k+1})_1, (\boldsymbol{X}_{k+1})_2, \cdots, (\boldsymbol{X}_{k+1})_n]$，按(9.9)式计算时，可先计算相邻两个迭代向量对应分量之比：

$$\begin{cases}
\dfrac{(\boldsymbol{X}_{k+1})_j}{(\boldsymbol{X}_k)_j} = (\lambda_1^{(k+1)})_j \qquad (j = 1,2,\cdots,n) \\
\dfrac{(\boldsymbol{X}_{k+2})_j}{(\boldsymbol{X}_{k+1})_j} = (\lambda_1^{(k+2)})_j \qquad (j = 1,2,\cdots,n)
\end{cases} \tag{9.10}$$

$$\cdots$$

再按

$$|(\lambda_1^{(k+1)})_j - (\lambda_1^{(k)})_j| < \varepsilon \qquad (j = 1,2,\cdots,n) \tag{9.11}$$

进行控制，在满足精度要求下，可取

$$\lambda_1 = \frac{1}{n}\sum_{j=1}^{n}(\lambda_1^{(k+1)})_j \tag{9.12}$$

② 为能求得按模最大的特征值 $\lambda_1$，要求 $\boldsymbol{X}_0$ 中的 $\alpha_1 \neq 0$。当 $\alpha_1=0$ 时，即 $\boldsymbol{X}_0$ 与 $\boldsymbol{V}_1$ 为直交的情况。一般情况下可设存在有 $i$，当 $\alpha_1=\alpha_2=\cdots=\alpha_{i-1}=0$ 而 $\alpha_i\neq 0$ 时就有

$$\boldsymbol{X}_0 = \alpha_i\boldsymbol{V}_i + \alpha_{i+1}\boldsymbol{V}_{i+1} + \cdots + \alpha_n\boldsymbol{V}_n$$

则用幂法得到的是 $\lambda_i$。但亦有如下的可能，即当 $\alpha_1=0$ 时，因舍入误差的原因，使

$$\boldsymbol{A}\boldsymbol{X}_0 = \lambda_1\varepsilon_1\boldsymbol{V}_1 + \alpha_2\lambda_2\boldsymbol{V}_2 + \cdots + \alpha_n\lambda_n\boldsymbol{V}_n$$

式中第一项虽小，但经很多次迭代后仍能得到稳定的 $\lambda_1$，如迭代次数甚大，可另选初始向量，即选取不与 $\boldsymbol{V}_1$ 直交的向量为初始向量，但在实践上无从判断，需要通过计算实践才能发现。

③ 式(9.10)中的比值趋于 $\lambda_1$ 的快慢取决于 $|\lambda_2/\lambda_1|$ 值的大小。

④ 式(9.10)中比值变化的一般特点是单调地变化；如果 $(\boldsymbol{X}_k)_j\ (k=0,1,2,\cdots)$ 的符号发生改变，一般也是正负相间地变化。

**例 9.1** 求 $\boldsymbol{A}$ 的最大特征值

$$\boldsymbol{A} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

**解** 取 $\boldsymbol{X}_0=[1,0,0]'$，按 $\boldsymbol{X}_{k+1}=\boldsymbol{A}\boldsymbol{X}_k$ 算得表 9.1。

**表 9.1**

| $k$ | 0 | 1 | … | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|
| $(\boldsymbol{X}_k)_1$ | 1 | 2 | | 16 016 | 54 320 | 184 736 | 629 280 |
| $(\boldsymbol{X}_k)_2$ | 0 | −1 | … | −22 288 | −76 096 | −259 808 | −887 040 |
| $(\boldsymbol{X}_k)_3$ | 0 | 0 | | 15 504 | 53 296 | 182 688 | 625 184 |

比值的变化如表 9.2 所示。

**表 9.2**

| $k$ | 9 | 10 | 11 |
|---|---|---|---|
| $(\boldsymbol{X}_{k+1})_1/(\boldsymbol{X}_k)_1$ | 3.39 | 3.40 | 3.41 |
| $(\boldsymbol{X}_{k+1})_2/(\boldsymbol{X}_k)_2$ | 3.41 | 3.41 | 3.41 |
| $(\boldsymbol{X}_{k+1})_3/(\boldsymbol{X}_k)_3$ | 3.44 | 3.42 | 3.42 |

**例 9.2** 求 $\boldsymbol{A}$ 的最大特征值

$$\boldsymbol{A} = \begin{bmatrix} -2 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -2 \end{bmatrix} \qquad \boldsymbol{X}_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

**解** 按 $\boldsymbol{X}_{k+1}=\boldsymbol{A}\boldsymbol{X}_k$ 计算得表 9.3。

表 9.3

| $k$ | 0 | 1 | 2 | 3 | 4 | … |
|---|---|---|---|---|---|---|
| $(X_k)_1$ | 1 | -2 | 5 | -14 | 42 | |
| $(X_k)_2$ | 0 | 1 | -4 | 14 | -48 | … |
| $(X_k)_3$ | 0 | 0 | 0 | -6 | 26 | |

由表值可见，$(X_k)_j$ 的符号随着 $k$ 的增长发生正负相间的变化。

⑤ 在计算过程中，可能出现绝对值过大的数（当 $|\lambda_1|>1$ 时）或过小的数（当 $|\lambda_1|<1$ 时），为避免这种情况的出现，实际计算时，可对每次迭代所求得的向量作归一化处理。因此，幂法实际使用的迭代公式是

$$\begin{cases} X_{k+1}=AY_k \\ Y_{k+1}=\dfrac{X_{k+1}}{\alpha}, \qquad \alpha=\max\limits_j|(X_{k+1})_j| \end{cases} \tag{9.13}$$

⑥ 因 $X_{k+1}\approx(\lambda_1^{k+1}\alpha_1)V_1$，所以可以取 $X_{k+1}$ 作为特征向量 $V_1$ 的近似，即

$$V_1\approx X_{k+1} \tag{9.14}$$

例如，例 9.1 中，与 $\lambda_1$ 对应的特征向量 $V_1$ 可取为 $X_{12}$ 得

$$V_1=[629280,-887040,625184]'$$

归一化后为

$$V_1=[0.709,-1,0.705]'$$

(2)$\lambda_1$ 是 $r$ 重根且 $|\lambda_1|>|\lambda_{r+1}|\geqslant\cdots\geqslant|\lambda_n|$ 情况

当 $k$ 充分大时，同法可得

$$X_k\approx\lambda_1^k[\alpha_1V_1+\alpha_2V_2+\cdots+\alpha_rV_r]$$

$$X_{k+1}\approx\lambda_1^{k+1}[\alpha_1V_1+\alpha_2V_2+\cdots+\alpha_rV_r]$$

$$\frac{X_{k+1}}{X_k}\approx\lambda_1 \tag{9.15}$$

或

$$\frac{(X_{k+1})_j}{(X_k)_j}\approx\lambda_1 \quad (j=1,2,\cdots,n) \tag{9.16}$$

$X_{k+1}$ 为 $\lambda_1$ 的一个特征向量，它随不同的 $X_0$ 而异。为了获取线性独立的特征向量，可利用伪随机数发生子程序随机地选取向量的分量形成线性独立的初始向量，分别地按幂法进行迭代，相应地就可获得几个线性独立的特征向量。

(3)$\lambda_1=-\lambda_2$；$|\lambda_1|>|\lambda_i|(i=3,4,\cdots,n)$ 情况

这时有

$$\begin{cases} X_k\approx\lambda_1^k[\alpha_1V_1+(-1)^k\alpha_2V_2] \\ X_{k+1}\approx\lambda_1^{k+1}[\alpha_1V_1+(-1)^{k+1}\alpha_2V_2] \\ X_{k+2}\approx\lambda_1^{k+2}[\alpha_1V_1+(-1)^{k+2}\alpha_2V_2] \end{cases} \tag{9.17}$$

当 $k$ 充分大时，迭代序列 $\{X_k\}$ 呈有规律的摆动，相邻向量的对应分量之比也总是波动不定，$X_{2k}$ 与 $X_{2k+2}$ 或 $X_{2k-1}$ 与 $X_{2k+1}$ 几乎仅差一个常数因子 $\lambda_1^2$。于是有

$$|\lambda_1|_j=\sqrt{\frac{(X_{k+2})_j}{(X_k)_j}} \qquad (j=1,2,\cdots,n) \tag{9.18}$$

由式(9.17)可以得到

$$\begin{cases} X_{k+1}+\lambda_1 X_k \approx 2\lambda_1^{k+1}\alpha_1 V_1 \\ X_{k+1}-\lambda_1 X_k \approx (-1)^{k+1}2\lambda_1^{k+1}\alpha_2 V_2 \end{cases} \tag{9.19}$$

因此,特征向量可取为

$$\begin{cases} V_1 \approx X_{k+1}+\lambda_1 X_k \\ V_2 \approx X_{k+1}-\lambda_1 X_k \end{cases} \tag{9.20}$$

**例 9.3** 按幂法计算下列矩阵

$$A=\begin{bmatrix} 5 & 4 & -4 \\ 11 & 8 & -8 \\ 13 & 13 & -12 \end{bmatrix}$$

按模最大的特征值和对应的特征向量。

**解** 取 $X_0=[1,0,0]'$,按 $X_{k+1}=AX_k$ 计算得表 9.4。

表 9.4

| $k$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| $(X_k)_1$<br>$(X_k)_2$<br>$(X_k)_3$ | 1<br>0<br>0 | 5<br>11<br>13 | 17<br>39<br>52 | 33<br>83<br>104 | 81<br>195<br>260 | 145<br>371<br>468 | 337<br>819<br>1 092 | 593<br>1 523<br>1 924 |
| $\frac{(X_{k+1})_1}{(X_k)_1}$ | 5 | 3.4 | 1.9 | 2.5 | 1.8 | 2.3 | 1.8 | |
| $\frac{(X_{k+2})_1}{(X_k)_1}$ | 17 | 6.6 | 4.8 | 4.4 | 4.1 | 4.1 | | |

$$\lambda_1^2 \approx 4.1$$

$$\lambda_1,\lambda_2 \approx \pm 2$$

$$V_1 \approx X_7+2X_6=\begin{bmatrix} 593 \\ 1\,523 \\ 1\,924 \end{bmatrix}+2\begin{bmatrix} 337 \\ 819 \\ 1\,092 \end{bmatrix}=\begin{bmatrix} 1\,267 \\ 3\,161 \\ 4\,108 \end{bmatrix}$$

$$V_2 \approx X_7-2X_6=\begin{bmatrix} 593 \\ 1\,523 \\ 1\,924 \end{bmatrix}-2\begin{bmatrix} 337 \\ 819 \\ 1\,092 \end{bmatrix}=\begin{bmatrix} -81 \\ -115 \\ -260 \end{bmatrix}$$

(4)$\lambda_1=\overline{\lambda}_2=\rho e^{i\theta}$, $|\lambda_1|=\rho>|\lambda_i|$ $(i=3,4,\cdots,n)$情况

因 $A$ 为实矩阵,当特征值为共轭复数时,其相应的特征向量也有共轭关系 $V_1=\overline{V}_2$。初始向量 $X_0$ 取实向量时,有

$$\begin{aligned} X_0 &= \alpha_1 V_1+\alpha_2 V_2+\alpha_3 V_3+\cdots+\alpha_n V_n \\ &= \alpha_1 V_1+\alpha_2 \overline{V}_1+\alpha_3 V_3+\cdots+\alpha_n V_n \end{aligned}$$

因 $X_0$ 为实向量,所以上式中 $\alpha_1$ 与 $\alpha_2$ 亦互为共轭量,即

$$X_0=\alpha_1 V_1+\overline{\alpha}_1\overline{V}_1+\alpha_3 V_3+\cdots+\alpha_n V_n \tag{9.21}$$

从而有

$$X_k=\lambda_1^k\alpha_1 V_1+\overline{\lambda}_1^k\overline{\alpha}_1\overline{V}_1+\lambda_3^k\alpha_3 V_3+\cdots+\lambda_n^k\alpha_n V_n$$

$$= \rho^k (e^{i\theta})^k \alpha_1 \boldsymbol{V}_1 + \rho^k (e^{-i\theta})^k \overline{\alpha}_1 \overline{\boldsymbol{V}}_1 + \lambda_3^k \alpha_3 \boldsymbol{V}_3 + \cdots + \lambda_n^k \alpha_n \boldsymbol{V}_n$$

$$\approx \rho^k [\alpha_1 e^{ik\theta} \boldsymbol{V}_1 + \overline{\alpha}_1 \, e^{-ik\theta} \overline{\boldsymbol{V}}_1] \tag{9.22}$$

$$(\boldsymbol{X}_k)_j \approx \rho^k [\alpha_1 e^{ik\theta} (\boldsymbol{V}_1)_j + \overline{\alpha}_1 e^{-ik\theta} (\boldsymbol{V}_1)_j] \tag{9.23}$$

令
$$\alpha_1 (\boldsymbol{V}_1)_j = R_j e^{i\varphi_j}$$

则
$$\overline{\alpha}_1 (\overline{\boldsymbol{V}}_1)_j = \overline{\alpha_1 (\boldsymbol{V}_1)_j} = R_j \, e^{-i\varphi_j}$$

$$(\boldsymbol{X}_k)_j \approx \rho^k [R_j (e^{i(\varphi_j + k\theta)} + e^{-i(\varphi_j + k\theta)})]$$

$$= 2R_j \rho^k \cos(k\theta + \varphi_j) \tag{9.24}$$

从式(9.24)还难于看出 $\boldsymbol{X}_k$ 的数值及符号上的变化有什么规律性，但由

$$\begin{cases} (\boldsymbol{X}_k)_j \approx 2R_j \rho^k \cos(k\theta + \varphi_j) \\ (\boldsymbol{X}_{k+1})_j \approx 2R_j \rho^{k+1} \cos[(k+1)\theta + \varphi_j] \\ (\boldsymbol{X}_{k+2})_j \approx 2R_j \rho^{k+2} \cos[(k+2)\theta + \varphi_j] \end{cases} \tag{9.25}$$

可得到
$$(\boldsymbol{X}_{k+2})_j - (\lambda_1 + \lambda_2)(\boldsymbol{X}_{k+1})_j + \lambda_1 \lambda_2 (\boldsymbol{X}_k)_j \approx 0 \quad (j = 1, 2, \cdots, n) \tag{9.26}$$

记 $-(\lambda_1 + \lambda_2) = p, \lambda_1 \lambda_2 = q$，可得

$$(\boldsymbol{X}_{k+2})_j + p(\boldsymbol{X}_{k+1})_j + q(\boldsymbol{X}_k)_j \approx 0 \tag{9.27}$$

对充分大的 $k$，式(9.27)中的≈以＝看待，就可得以下方程组

$$(\boldsymbol{X}_{k+1})_j p + (\boldsymbol{X}_k)_j q = -(\boldsymbol{X}_{k+2})_j \quad (j = 1, 2, \cdots, n) \tag{9.28}$$

或改写为

$$\begin{cases} a_{11} p + a_{12} q = b_1 \\ a_{21} p + a_{22} q = b_2 \\ \cdots \\ a_{n1} p + a_{n2} q = b_n \end{cases} \tag{9.29}$$

式(9.29)用矩阵表为

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \cdots & \\ a_{n1} & a_{n2} \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_n \end{bmatrix} \tag{9.30}$$

或 $\boldsymbol{AX} = \boldsymbol{B}$ (9.31)

其中
$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \cdots & \cdots \\ a_{n1} & a_{n2} \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} p \\ q \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_n \end{bmatrix} \tag{9.32}$$

对式(9.31)采用最小平方法求解得以下正规方程组

$$\boldsymbol{A}'\boldsymbol{AX} = \boldsymbol{A}'\boldsymbol{B} \tag{9.33}$$

在解得 $p, q$ 值的基础上，就有

$$\begin{cases} -(\lambda_1 + \lambda_2) = p \\ \lambda_1 \lambda_2 = q \end{cases}$$

即 $\lambda_1 \lambda_2$ 是方程 $\lambda^2 + p\lambda + q = 0$ 的根，从而有

$$\lambda_1 = -\frac{p}{2} + \sqrt{\frac{p^2}{4} - q}, \quad \lambda_2 = -\frac{p}{2} - \sqrt{\frac{p^2}{4} - q} \tag{9.34}$$

又因为 $\quad X_{k+1}-\lambda_2X_k\approx[\lambda_1^{k+1}\alpha_1V_1+\lambda_2^{k+1}\alpha_2V_2]-\lambda_2[\lambda_1^k\alpha_1V_1+\lambda_2^k\alpha_2V_2]=\lambda_1^k(\lambda_1-\lambda_2)\alpha_1V_1$

及 $\quad X_{k+1}-\lambda_1X_k\approx[\lambda_1^{k+1}\alpha_1V_1+\lambda_2^{k+1}\alpha_2V_2]-\lambda_1[\lambda_1^k\alpha_1V_1+\lambda_2^k\alpha_2V_2]=\lambda_2^k(\lambda_2-\lambda_1)\alpha_2V_2$

因而可取

$$\begin{cases}V_1\approx X_{k+1}-\lambda_2X_k\\V_2\approx X_{k+1}-\lambda_1X_k\end{cases}\tag{9.35}$$

在实际计算中,需要根据迭代序列数值变化的特性来判定属于何种情况,然后才能做出相应的处理。如果迭代序列的比值$(X_{k+1})_j/(X_k)_j$逐渐趋近一个定值,那么这个定值就是$A$的绝对值最大的特征值$\lambda_1$的一个近似值。倘若上述比值总是波动不定,就说明$A$至少有两个绝对值相等但自身并不相等的特征值$\lambda_1$与$\lambda_2$,自然这个论断还蕴含有各种可能性,以$|\lambda_1|=|\lambda_2|>|\lambda_i|(i\geqslant3)$的情况来讲,就有以下两种可能。

① $\lambda_1=-\lambda_2$都是实数,那么$(X_{k+2})_j/(X_k)_j$将逐渐趋近$\lambda_1^2$。

② 如果下式

$$(X_{k+2})_j+p(X_{k+1})_j+q(X_k)_j$$

有较小的数值时,则可考虑为属于$\lambda_1=\overline{\lambda_2}$的情况。若均不属于以上的更为复杂的情况时,使用幂法可能失败而难于获得所需的结果,可考虑改用其他的方法。

### 1.2 反幂法

反幂法是计算矩阵按模最小特征值及特征向量的方法。设$|\lambda_i|\neq0(i=1,2,\cdots,n)$,于是有$|A|=\lambda_1\lambda_2\cdots\lambda_n\neq0$,所以$A^{-1}$存在。由$AX=\lambda X$可得

$$A^{-1}X=\frac{1}{\lambda}X=\Lambda X\tag{9.36}$$

上式表明,$A^{-1}$的特征值为$\Lambda=1/\lambda$。于是$A$的按模最小的特征值$\lambda_n$正是$A^{-1}$按模最大的特征值$1/\lambda_n$,而相应的特征向量不变。因此,若对矩阵$A^{-1}$用幂法,算出其按模最大的特征值,则其倒数恰为$A$的按模最小的特征值。这就是反幂法的基本思想。

用幂法求$A^{-1}$按模最大特征值仍用归一化方法运算,取$X_0=Y_0$,然后作迭代

$$\begin{cases}X_{k+1}=A^{-1}Y_k\\Y_{k+1}=\dfrac{X_{k+1}}{\alpha},\quad\alpha=\max\limits_j|(X_{k+1})_j|\end{cases}\tag{9.37}$$

为了避免矩阵求逆,式(9.37)的迭代式可通过解方程组

$$AX_{k+1}=Y_k\tag{9.38}$$

得到$X_{k+1}$。

反幂法还可用于求与$\mu$值最靠近的特征值和特征向量。设$\lambda_k$与$\mu$值最靠近,则有

$$|\lambda_k-\mu|<|\lambda_i|\qquad(i\neq k)\tag{9.39}$$

由$AX=\lambda_kX$两边减去$\mu X$得

$$(A-\mu I)X=(\lambda_k-\mu)X\tag{9.40}$$

上式表明,$\lambda_k-\mu$是矩阵$A-\mu I$的特征值且是按模最小的特征值。对于式(9.40)应用反幂法求得$(A-\mu I)^{-1}$按模最大的特征值$\Lambda=1/(\lambda_k-\mu)$,则与$\mu$值最靠近的特征值为$\lambda_k=\mu+\dfrac{1}{\Lambda}$。

# §2　正交变换矩阵

作为求解矩阵特征值的另一类方法，就是利用矩阵变换的方法求矩阵的特征值。其基本思想是根据相似矩阵必有相同的特征谱这一性质，把矩阵化为形式简单的矩阵，使新矩阵的特征值便于计算。其中正交矩阵将是作变换所用的一类矩阵，它们是计算特征值问题及其他矩阵计算问题的有力工具。

## 2.1　正交与正交矩阵

### 2.1.1　正交的概念

**定义 9.1**　设有非零向量 $\boldsymbol{X},\boldsymbol{Y}$，如果向量内积 $(\boldsymbol{X},\boldsymbol{Y})=0$，则称 $\boldsymbol{X},\boldsymbol{Y}$ 正交或互相垂直。

在二或三维情况下，向量内积的计算公式可以表示为

$$(\boldsymbol{X},\boldsymbol{Y}) = |\boldsymbol{X}|\,|\boldsymbol{Y}|\cos\theta \tag{9.41}$$

在二维或多维情况下，当每个向量的坐标已知时，设

$$\boldsymbol{X}=\begin{bmatrix}x_1\\x_2\\\cdots\\x_n\end{bmatrix},\qquad \boldsymbol{Y}=\begin{bmatrix}y_1\\y_2\\\cdots\\y_n\end{bmatrix}$$

则向量内积的计算公式可表示为

$$(\boldsymbol{X},\boldsymbol{Y}) = \boldsymbol{X}'\boldsymbol{Y} = [x_1,x_2,\cdots,x_n]\begin{bmatrix}y_1\\y_2\\\cdots\\y_n\end{bmatrix} = \sum_{i=1}^{n}x_iy_i \tag{9.42}$$

**定义 9.2**　$\mathbf{R}^n$ 中一组非零向量 $\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_m,(m\leqslant n)$ 称为正交的，若其中任何两个向量互相正交，即满足

$$(\boldsymbol{X}_j,\boldsymbol{X}_k)=0\qquad (j\neq k) \tag{9.43}$$

则称 $\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_m$ 为正交向量组。易证，正交向量必线性独立。

### 2.1.2　正交矩阵及其性质

**定义 9.3**　在 $\mathbf{R}^{n\times n}$ 中，若矩阵 $\boldsymbol{A}$ 的列由 $n$ 个单位长度的正交向量 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n$，所组成，则称矩阵 $\boldsymbol{A}$ 为正交矩阵。它有以下性质：

① 正交矩阵所体现的正交变换不改变被变换向量的长度和它们间的夹角，即若 $\boldsymbol{A}$ 为正交矩阵，$\boldsymbol{X}$、$\boldsymbol{Y}$ 为向量，则有 $\|\boldsymbol{X}\|=\|\boldsymbol{AX}\|$；由于 $(\boldsymbol{X},\boldsymbol{Y})=(\boldsymbol{AX},\boldsymbol{AY})$，所以有 $\cos(\boldsymbol{X},\boldsymbol{Y})=\cos(\boldsymbol{AX},\boldsymbol{AY})$。

② $\boldsymbol{A}'\boldsymbol{A}=\boldsymbol{I}$。　(9.44)

因 $\boldsymbol{A}$ 的列向量满足 $(\boldsymbol{a}_i,\boldsymbol{a}_j)=0(i\neq j)$ 和 $(\boldsymbol{a}_i,\boldsymbol{a}_j)=1(i=j)$，由此即得(9.44)式。

③ $\boldsymbol{A}'=\boldsymbol{A}^{-1}$。　(9.45)

由 $\boldsymbol{A}'\boldsymbol{A}=\boldsymbol{I}$ 及 $\boldsymbol{A}^{-1}\boldsymbol{A}=\boldsymbol{I}$ 即知(9.45)式成立。

④ $\boldsymbol{AA}'=\boldsymbol{I}$。　(9.46)

因 $\boldsymbol{AA}'=\boldsymbol{AA}^{-1}=\boldsymbol{I}$，可见 $\boldsymbol{A}$ 的行向量亦构成正交向量组，即 $\boldsymbol{A}'$ 也是正交矩阵。

综上所述,若 $\boldsymbol{A}$ 为正交矩阵,则以下命题等价

$$\boldsymbol{A}'\boldsymbol{A}=\boldsymbol{I}\Leftrightarrow\boldsymbol{A}\boldsymbol{A}'=\boldsymbol{I}\Leftrightarrow\boldsymbol{A}'=\boldsymbol{A}^{-1} \tag{9.47}$$

⑤ 正交矩阵之积仍为正交矩阵。

事实上,若 $\boldsymbol{A},\boldsymbol{B}$ 为正交矩阵,则有

$$(\boldsymbol{A}\boldsymbol{B})'=\boldsymbol{B}'\boldsymbol{A}'=\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}=(\boldsymbol{A}\boldsymbol{B})^{-1}$$

由式(9.47)知 $\boldsymbol{A}\boldsymbol{B}$ 亦为正交矩阵。

⑥ 正交矩阵的逆矩阵亦为正交矩阵。

因$(\boldsymbol{A}^{-1})'=(\boldsymbol{A}')'=(\boldsymbol{A}^{-1})^{-1}$,知 $\boldsymbol{A}^{-1}$为正交矩阵。

⑦ 正交矩阵行列式的值等于±1。

由 $\boldsymbol{A}\boldsymbol{A}'=\boldsymbol{I}$ 得$|\boldsymbol{A}|^2=1$ 或$|\boldsymbol{A}|=\pm1$。

⑧ 用正交矩阵左乘或右乘矩阵 $\boldsymbol{A}$,其 $F$——范数不变。

**定义 9.4** 矩阵 $\boldsymbol{A}$ 的对角线元素之和称为 $\boldsymbol{A}$ 的迹。记为 $\mathrm{tr}(\boldsymbol{A})=\sum_{i=1}^{n}a_{ii}$。

设 $\boldsymbol{Q}$ 与 $\boldsymbol{R}$ 为正交矩阵,则有

$\|\boldsymbol{A}\|_F^2=\sum_{i,j=1}^{n}a_{ij}^2=\mathrm{tr}(\boldsymbol{A}'\boldsymbol{A})=\mathrm{tr}(\boldsymbol{A}'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{A})=\mathrm{tr}((\boldsymbol{Q}\boldsymbol{A})'(\boldsymbol{Q}\boldsymbol{A}))=\|\boldsymbol{Q}\|_F^2$。因 $\boldsymbol{R}$ 为正交矩阵,所以 $\boldsymbol{R}'$也是正交矩阵,从而又可得

$$\|\boldsymbol{A}\|_F^2=\|\boldsymbol{A}'\|_F^2=\|\boldsymbol{R}'\boldsymbol{A}'\|_F^2=\|(\boldsymbol{A}\boldsymbol{R})'\|_F^2=\|\boldsymbol{A}\boldsymbol{R}\|_F^2$$

同样有

$$\|\boldsymbol{Q}\boldsymbol{A}\boldsymbol{R}\|_F^2=\|\boldsymbol{A}\|_F^2。 \tag{9.48}$$

## 2.2 旋转变换矩阵

在平面解析几何中曾学过某点 $M$ 的坐标在新、旧坐标系中的旋转变换公式(图 9.1)为

$$\begin{cases}x_1=x_2\cos\varphi-y_2\sin\varphi\\ y_1=x_2\sin\varphi+y_2\cos\varphi\end{cases} \tag{9.49}$$

其变换矩阵为

$$\boldsymbol{V}=\begin{bmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{bmatrix} \tag{9.50}$$

称为旋转变换矩阵。

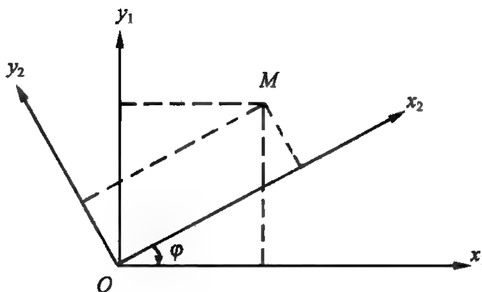



图 9.1

在 $n$ 维空间中，仿此可设计如下旋转变换矩阵

$$V_{ij}(\varphi)=\begin{bmatrix} 1 & & & \overset{i列}{} & & \overset{j列}{} & & & \\ & \ddots & & | & & | & & & \\ & & 1 & & & & & & \\ - & - & - & \cos\varphi & - & -\sin\varphi & - & - & i行 \\ & & & | & 1 & | & & & \\ & & & | & \ddots & | & & & \\ & & & | & & 1 & & & \\ - & - & - & \sin\varphi & - & \cos\varphi & - & - & j行 \\ & & & | & & | & 1 & & \\ & & & | & & | & & \ddots & \\ & & & | & & | & & & 1 \end{bmatrix} \tag{9.51}$$

它在 $n$ 维空间中将相互正交的两个坐标轴 $Ox_i$、$Ox_j$ 在其所决定的平面上旋转了一个角度 $\varphi$，而保持正交坐标系的其他轴不动。矩阵 $\boldsymbol{V}_{ij}(\varphi)$ 称为 $n$ 维空间内的平面旋转变换矩阵，简称旋转变换矩阵。容易验证 $\boldsymbol{V}_{ij}(\varphi)$ 是正交矩阵。如果用 $\boldsymbol{V}_{ij}(\varphi)$ 左乘矩阵 $\boldsymbol{A}$，即

$$\widetilde{\boldsymbol{A}}=\boldsymbol{V}_{ij}(\varphi)\boldsymbol{A} \tag{9.52}$$

则 $\widetilde{\boldsymbol{A}}$ 的元素为

$$\begin{cases}\tilde{a}_{kl}=a_{kl} & (k\neq i,j)\\ \tilde{a}_{il}=a_{il}\cos\varphi-a_{jl}\sin\varphi & \\ \tilde{a}_{jl}=a_{il}\sin\varphi+a_{jl}\cos\varphi & (l=1,2,\cdots,n)\end{cases} \tag{9.53}$$

可见 $\widetilde{\boldsymbol{A}}$ 与 $\boldsymbol{A}$ 仅 $i$ 行与 $j$ 行的元素不同。从式(9.53)还可以看出，适当地选择 $\varphi$，便可使 $\widetilde{\boldsymbol{A}}$ 的第 $j$ 行的任意取定的某一个元素 $\tilde{a}_{jl}=0(l=1,2,\cdots,n)$，特别取 $l=i$，并令

$$\begin{cases}\sin\varphi=\dfrac{-a_{ji}}{\sqrt{a_{ii}^2+a_{ji}^2}},\quad \cos\varphi=\dfrac{a_{ii}}{\sqrt{a_{ii}^2+a_{ji}^2}} & (\sqrt{a_{ii}^2+a_{ji}^2}\neq 0)\\ \sin\varphi=0,\quad \cos\varphi=1 & (\sqrt{a_{ii}^2+a_{ji}^2}=0)\end{cases} \tag{9.54}$$

则可使 $\tilde{a}_{ji}=0$，即按上式取定 $\varphi$ 时，用 $\boldsymbol{V}_{ij}(\varphi)$ 左乘矩阵 $\boldsymbol{A}$ 后可将 $a_{ji}$ 转化为 0。同样，用 $\boldsymbol{V}_{ij}'(\varphi)$ 左乘矩阵 $\boldsymbol{A}$，即

$$\hat{\boldsymbol{A}}=\boldsymbol{V}_{ij}'(\varphi)\boldsymbol{A} \tag{9.55}$$

则 $\hat{\boldsymbol{A}}$ 与 $\boldsymbol{A}$ 的元素亦仅在 $i$ 行，$j$ 行上不同。同法可以验证，若用 $\boldsymbol{V}_{ij}(\varphi)$ 或 $\boldsymbol{V}_{ij}'(\varphi)$ 右乘矩阵 $\boldsymbol{A}$，则变换后的矩阵与 $\boldsymbol{A}$ 仅 $i$ 列、$j$ 列上的元素不同。因此，若对 $\hat{\boldsymbol{A}}$ 再右乘以 $\boldsymbol{V}_{ij}(\psi)$，则变换后的矩阵与 $\boldsymbol{A}$ 必在 $i$ 行、$j$ 行与 $i$ 列、$j$ 列上的元素不同。记

$$\boldsymbol{A}^{(1)}=\hat{\boldsymbol{A}}\boldsymbol{V}_{ij}(\psi)=\boldsymbol{V}_{ij}'(\varphi)\boldsymbol{A}\boldsymbol{V}_{ij}(\psi) \tag{9.56}$$

则 $\boldsymbol{A}^{(1)}$ 的元素为

$$\begin{cases}a_{kl}^{(1)}=a_{kl} \qquad (k\neq i,j;l\neq i,j)\\ a_{il}^{(1)}=a_{il}\cos\varphi+a_{jl}\sin\varphi\\ a_{jl}^{(1)}=-a_{il}\sin\varphi+a_{jl}\cos\varphi\\ a_{li}^{(1)}=a_{li}\cos\psi+a_{lj}\sin\psi\\ a_{lj}^{(1)}=-a_{li}\sin\psi+a_{lj}\cos\psi\\ a_{ii}^{(1)}=(a_{ii}\cos\varphi+a_{ji}\sin\varphi)\cos\psi+(a_{ij}\cos\varphi+a_{jj}\sin\varphi)\sin\psi\\ a_{jj}^{(1)}=-(-a_{ii}\sin\varphi+a_{ji}\cos\varphi)\sin\psi+(-a_{ij}\sin\varphi+a_{jj}\cos\varphi)\cos\psi\\ a_{ij}^{(1)}=-(a_{ii}\cos\varphi+a_{ji}\sin\varphi)\sin\psi+(a_{ij}\cos\varphi+a_{jj}\sin\varphi)\cos\psi\\ a_{ji}^{(1)}=(-a_{ii}\sin\varphi+a_{ji}\cos\varphi)\cos\psi+(-a_{ij}\sin\varphi+a_{jj}\cos\varphi)\sin\psi\end{cases} \tag{9.57}$$

计算一下矩阵 $\boldsymbol{A}^{(1)}$ 非对角线元素的平方和,可以得到

$$\sum_{k\neq l}[a_{kl}^{(1)}]^2=\sum_{k\neq l}a_{kl}^2-(a_{ij}^2+a_{ji}^2)+[(-a_{ii}\sin\varphi+a_{ji}\cos\varphi)\cos\psi+(-a_{ij}\sin\varphi+a_{jj}\cos\varphi)\sin\psi]^2+[-(a_{ii}\cos\varphi+a_{ji}\sin\varphi)\sin\psi+(a_{ij}\cos\varphi+a_{jj}\sin\varphi)\cos\psi]^2 \tag{9.58}$$

如果矩阵 $\boldsymbol{A}$ 是对称的,并令 $\psi=\varphi$,则得相似矩阵 $\boldsymbol{A}^{(1)}$,且

$$(\boldsymbol{A}^{(1)})'=(\boldsymbol{V}_{ij}'(\varphi)\boldsymbol{A}\boldsymbol{V}_{ij}(\varphi))'=\boldsymbol{V}_{ij}'(\varphi)\boldsymbol{A}'(\boldsymbol{V}_{ij}'(\varphi))'=\boldsymbol{V}_{ij}'(\varphi)\boldsymbol{A}'\boldsymbol{V}_{ij}=\boldsymbol{A}^{(1)}$$

即 $\boldsymbol{A}^{(1)}$ 也是对称矩阵,并有如下公式成立

$$\begin{cases}a_{ij}^{(1)}=a_{ji}^{(1)}=-(a_{ii}\cos\varphi+a_{ji}\sin\varphi)\sin\varphi+(a_{ij}\cos\varphi+a_{jj}\sin\varphi)\cos\varphi\\ \qquad\qquad =a_{ij}(\cos^2\varphi-\sin^2\varphi)+(a_{jj}-a_{ii})\sin\varphi\cos\varphi\\ \sum\limits_{k\neq l}[a_{kl}^{(1)}]^2=\sum\limits_{k\neq l}a_{kl}^2-2a_{ij}^2+\frac{1}{2}[(a_{jj}-a_{ii})\sin2\varphi+2a_{ij}\cos2\varphi]^2\end{cases} \tag{9.59}$$

适当选取 $\varphi$ 角,使 $a_{ij}^{(1)}=a_{ji}^{(1)}=0$,即 $\varphi$ 应满足

$$a_{ij}\cos2\varphi+\frac{1}{2}(a_{jj}-a_{ii})\sin2\varphi=0$$

解得

$$\tan2\varphi=\frac{2a_{ij}}{a_{ii}-a_{jj}}\qquad\left(\text{或 } \mathrm{ctan}2\varphi=\frac{a_{ii}-a_{jj}}{2a_{ij}}\right) \tag{9.60}$$

按式(9.60)选定 $\varphi$ 角后,则式(9.59)化为

$$\begin{cases}a_{ij}^{(1)}=a_{ji}^{(1)}=0\\ \sum\limits_{k\neq l}[a_{kl}^{(1)}]^2=\sum\limits_{k\neq l}a_{kl}^2-2a_{ij}^2\end{cases} \tag{9.61}$$

从式(9.61)可见,只要按式(9.60)取定 $\varphi$,就可建立 $\boldsymbol{V}_{ij}(\varphi)$,对 $\boldsymbol{A}$ 作正交相似变换后,所得 $\boldsymbol{A}^{(1)}$,其 $a_{ij}^{(1)}=a_{ji}^{(1)}=0$,即非对角元的平方和的数值减少了 $2a_{ij}^2$;又据式(9.48)知 $\|\boldsymbol{A}^{(1)}\|_F^2=\|\boldsymbol{A}\|_F^2$,即 $\boldsymbol{A}^{(1)}$ 与 $\boldsymbol{A}$ 的元素之平方和总值不变,因此 $\boldsymbol{A}^{(1)}$ 的对角线元素的平方和的数值必增加了 $2a_{ij}^2$。

## 2.3 豪斯荷尔德矩阵

**定义 9.5** 设向量 $w$ 是 $n$ 维实向量,满足 $\|w\|_2=1$,则称矩阵

$$\boldsymbol{H}=\boldsymbol{I}-2ww' \tag{9.62}$$

为豪斯荷尔德矩阵或 $\boldsymbol{H}$ 矩阵,即

$$\boldsymbol{H}=\boldsymbol{I}-2ww'=\begin{bmatrix}1 & & 0\\ & \ddots & \\ 0 & & 1\end{bmatrix}-2\begin{bmatrix}w_1\\ w_2\\ \cdots\\ w_n\end{bmatrix}[w_1w_2\cdots w_n]$$

$$=\begin{bmatrix}1-2w_1^2 & -2w_1w_2\cdots & -2w_1w_n\\ -2w_2w_1 & 1-2w_2^2\cdots & -2w_2w_n\\ \cdots & & \\ -2w_nw_1 & -2w_nw_2\cdots & 1-2w_n^2\end{bmatrix}$$

**2.3.1　$\boldsymbol{H}$ 矩阵的性质**

① $\boldsymbol{H}$ 矩阵为对称矩称。

事实上，$\boldsymbol{H}'=(\boldsymbol{I}-2ww')'=\boldsymbol{I}-2(ww')'=\boldsymbol{I}-2ww'=\boldsymbol{H}$。

② $\boldsymbol{H}$ 矩阵是正交矩阵。

事实上，
$$\begin{aligned}\boldsymbol{H}'\boldsymbol{H} &=(\boldsymbol{I}-2ww')(\boldsymbol{I}-2ww')\\&=\boldsymbol{I}-4ww'+4(ww')(ww')\\&=\boldsymbol{I}-4ww'+4w(w'w)w'\end{aligned}$$

因 $w'w=\|w\|_2^2=1$，所以

$$\boldsymbol{H}'\boldsymbol{H}=\boldsymbol{I}-4ww'+4ww'=\boldsymbol{I}$$

即 $\boldsymbol{H}$ 矩阵为正交矩阵。因此有 $\boldsymbol{H}^{-1}=\boldsymbol{H}'=\boldsymbol{H}$，即 $\boldsymbol{H}$ 的逆矩阵就是 $\boldsymbol{H}$ 本身，因而用它来作相似变换非常方便。

③ 镜面反射特性。

下面考察 $\boldsymbol{H}$ 矩阵变换的几何意义。考虑以 $w$ 为法向量的 $n-1$ 维子空间（可理解为一个超平面 $S$，如图 9.2 所示）

$$S=\{\boldsymbol{X}\mid_{w'\boldsymbol{X}_1=0}\}$$

设任意向量 $\boldsymbol{X}\in\mathbf{R}^n$，作正交分解 $\boldsymbol{X}=\boldsymbol{X}_1+\boldsymbol{X}_2$，其中 $\boldsymbol{X}_1\in S$，$\boldsymbol{X}_2\perp S$（图 9.2）。因 $\boldsymbol{X}_2\perp S$，所以 $\boldsymbol{X}_2$ 平行于 $w$，可令 $\boldsymbol{X}_2=cw$（$c$ 为常数），则有

$$\begin{aligned}\boldsymbol{HX}&=(\boldsymbol{I}-2ww')(\boldsymbol{X}_1+\boldsymbol{X}_2)\\&=\boldsymbol{X}_1-2w(w'\boldsymbol{X}_1)+\boldsymbol{X}_2-2ww'\boldsymbol{X}_2\\&=\boldsymbol{X}_1+\boldsymbol{X}_2-2cw(w'w)\qquad(\text{因 } w'\boldsymbol{X}_1=0)\\&=\boldsymbol{X}_1+\boldsymbol{X}_2-2cw\\&=\boldsymbol{X}_1+\boldsymbol{X}_2-2\boldsymbol{X}_2=\boldsymbol{X}_1-\boldsymbol{X}_2=\boldsymbol{Y}\end{aligned}\tag{9.63}$$

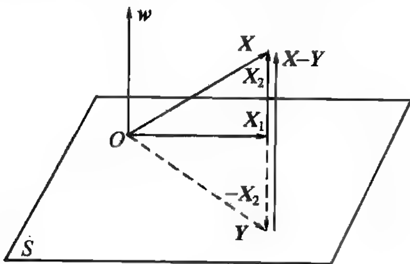



图 9.2

由图 9.2 可见，$\boldsymbol{Y}$ 与 $\boldsymbol{X}$ 相对于平面 $S$ 对称，称为镜面反射。因此 $\boldsymbol{H}$ 矩阵也叫镜面反射矩阵或初等反射矩阵。显见

$$\|\boldsymbol{HX}\|_2=\|\boldsymbol{Y}\|_2=\|\boldsymbol{X}\|_2\tag{9.64}$$

即在 $\boldsymbol{H}$ 矩阵变换下，其向量的长度保持不变。

**2.3.2　关于 $\boldsymbol{H}$ 矩阵的构造方法**

构造 $\boldsymbol{H}$ 矩阵主要是确定向量 $w$，但这并不困难，对任意非零向量 $\boldsymbol{V}\in\mathbf{R}^n$，只要令

$$w=\frac{\boldsymbol{V}}{\|\boldsymbol{V}\|_2}\tag{9.65}$$

则 $\|w\|_2=\left\|\frac{V}{\|V\|_2}\right\|_2=1$，于是

$$H=I-2ww'=I-2\frac{V}{\|V\|_2}\cdot\frac{V'}{\|V\|_2}$$

$$=I-2\frac{VV'}{\|V\|_2^2}=I-\frac{VV'}{\frac{1}{2}\|V\|_2^2}$$

$$=I-\beta^{-1}VV' \tag{9.66}$$

式中，$\beta=\frac{1}{2}\|V\|_2^2$ (9.67)

#### 2.3.3 关于 $HX=Y$ 的 $H$ 矩阵构造方法

如果我们给定了空间二个向量 $X$ 和 $Y$，满足 $\|X\|_2=\|Y\|_2$，则可找到 $H$ 矩阵，使

$$HX=Y \tag{9.68}$$

要确定这样的 $H$ 矩阵，关键是要确定对应的 $w$。从图 9.2 看出 $w$ 平行于 $(X-Y)$，所以我们应取

$$w=\frac{X-Y}{\|X-Y\|_2} \tag{9.69}$$

不难验证，这样确定的 $H$ 矩阵，式(9.68)成立。

#### 2.3.4 关于 $HX=(*,0,\cdots,0)'$ 的 $H$ 矩阵构造方法

对不等于零的空间向量 $X$，要求这样的 $H$ 矩阵，它可以将 $X$ 除第一个分量外的其余分量化为零。这个要求用公式来表示即为

$$HX=k_1e_1 \tag{9.70}$$

其中 $k_1$ 为常数，$e_1=(1,0,\cdots,0)'$。由(9.68)式知，只需取 $Y=k_1e_1$ 即可。这时有

$$\|k_1e_1\|_2=\|X\|_2$$

可取 $k_1=\pm\|X\|_2$，　$Y=\pm\|X\|_2e_1$ (9.71)

代入式(9.69)得

$$w=\frac{X-(\pm\|X\|_2)e_1}{\|X-(\pm\|X\|_2)e_1\|_2} \tag{9.72}$$

设 $X=(x_1,x_2,\cdots,x_n)'$，为了使 $X-k_1e_1$ 计算时避免由于相近数相减而导致损失有效数位，可取

$$k_1=-\operatorname{sgn}(x_1)\|X\|_2=-\sigma_1 \tag{9.73}$$

则得 $w=\frac{X+\operatorname{sgn}(x_1)\|X\|_2e_1}{\|X+\operatorname{sgn}(x_1)\|X\|_2e_1\|_2}=\frac{X+\sigma_1e_1}{\|X+\sigma_1e_1\|_2}=\frac{V}{\|V\|_2}$ (9.74)

其中 $\sigma_1=\operatorname{sgn}(x_1)\|X\|_2$

$$V=X+\sigma_1e_1=(x_1+\operatorname{sgn}(x_1)\|X\|_2,x_2,\cdots,x_n)'$$
$$=(x_1+\sigma_1,x_2,\cdots,x_n)'$$

则 $\beta=\frac{1}{2}\|V\|_2^2=\frac{1}{2}V'V=\frac{1}{2}(X+\sigma_1e_1)'(X+\sigma_1e_1)$

$$=\frac{1}{2}[X'X+\sigma_1X'e_1+\sigma_1e'X+\sigma_1^2e_1'e_1]$$

$$=\frac{1}{2}[\|X\|_2^2+\sigma_1x_1+\sigma_1x_1+\sigma_1^2]$$

$$=\frac{1}{2}[\|X\|_2^2+2\sigma_1x_1+\sigma_1^2]$$

$$=\frac{1}{2}[\|X\|_2^2+2\text{sgn}(x_1)\|X\|_2\cdot x_1+(\text{sgn}(x_1)\|X\|_2)^2]$$

$$=\frac{1}{2}[2\|X\|_2^2+2\|X\|_2|x_1|]=\|X\|_2(\|X\|_2+|x_1|)$$

综上，可以得到计算 $H$ 矩阵的步骤如下。

① 计算

$$\sigma_1=\text{sgn}(x_1)\|X\|_2 \tag{9.75}$$

② 计算

$$V=(x_1+\sigma_1,x_2,\cdots,x_n)' \tag{9.76}$$

③ 计算

$$\beta=\|X\|_2(\|X\|_2+|x_1|) \tag{9.77}$$

④ 写出

$$H=I-\beta^{-1}VV' \tag{9.78}$$

容易验证 $HX=(-\sigma_1,0,\cdots,0)'$

**例 9.4**　设有向量 $X=(2,1,2)'$，构造 $H$ 矩阵，使 $HX=k_1e_1$，$e_1=(1,0,0)'$。

**解**　按式(9.75)～式(9.78)计算如下：

① $\sigma_1=\text{sgn}(x_1)\|X\|_2=+\sqrt{2^2+1^2+2^2}=3$

② $V=X+\sigma_1e_1=\begin{bmatrix}2\\1\\2\end{bmatrix}+3\begin{bmatrix}1\\0\\0\end{bmatrix}=\begin{bmatrix}5\\1\\2\end{bmatrix}$

③ $\beta=3(3+2)=15$

④ 计算

$$H=I-\beta^{-1}VV'=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}-\frac{1}{15}\begin{bmatrix}5\\1\\2\end{bmatrix}[5\quad 1\quad 2]$$

$$=\frac{1}{15}\begin{bmatrix}-10&-5&-10\\-5&14&-2\\-10&-2&11\end{bmatrix}$$

容易验证

$$HX=\begin{bmatrix}-3\\0\\0\end{bmatrix}$$

**2.3.5　关于 $HX=(*,*,\cdots,*,0,\cdots,0)'$ 的 $H$ 矩阵构造方法**

对 $\mathbf{R}^n$ 中的向量 $X=(x_1,\cdots,x_{j-1},x_j,\cdots,x_n)$，要求构造一个 $H$ 矩阵，使 $HX$ 第 $(j+1)\sim n$ 个分量均为 0。记

$$\tilde{X}=(x_j,x_{j+1},\cdots,x_n)'$$

$$\sigma_j=\text{sgn}(x_j)\|\tilde{X}\|_2$$

$$\tilde{V}=(x_j+\sigma_j,x_{j+1},\cdots,x_n)'$$

$$\beta = \frac{1}{2}\|\widetilde{V}\|_2^2 = \|\widetilde{X}\|_2(\|\widetilde{X}\|_2 + |x_j|)$$

$$\widetilde{H} = I_{n-(j-1)} - \beta^{-1}\widetilde{V}\widetilde{V}' \tag{9.79}$$

这里 $\widetilde{H}$ 为$n-(j-1)$阶 $H$ 矩阵。今取

$$V = (0,\cdots,0,x_j+\sigma_j,x_{j+1},\cdots,x_n)' \tag{9.80}$$

由此确定出以下矩阵

$$H = I - \beta^{-1}VV' = \begin{bmatrix} I_{j-1} & 0 \\ 0 & \widetilde{H} \end{bmatrix} \tag{9.81}$$

容易验证 $HX=X-V=(x_1,\cdots,x_{j-1},-\sigma_j,0,\cdots0)'$

式(9.81)的 $H$ 矩阵尚有以下有用特性：用它去左乘一个矩阵 $A$，将保持 $A$ 的前 $j-1$ 行不变；而用它去右乘一个矩阵 $A$，将保持 $A$ 的前 $j-1$ 列不变。

### 2.3.6 关于 $HX=(*,\cdots,*,0,\cdots,0,*,\cdots,*)'$的 $H$ 矩阵构造方法

类似地也可以将 $X$ 变换为 $x_j$ 与 $x_k$ 之间的所有分量为 0 的向量。设 $1\leqslant j<k\leqslant n$，$X\in\mathbf{R}^n$ 为非零向量，记

$$\begin{cases} X = (x_1,\cdots,x_{j-1},x_j,\cdots,x_k,x_{k+1},\cdots,x_n)' \\ \widetilde{X} = (x_j,\cdots,x_k)' \\ \sigma_j = \mathrm{sgn}(x_j)\|\widetilde{X}\|_2 \\ V = (0,\cdots,0,x_j+\sigma_j,x_{j+1},\cdots,x_k,0,\cdots,0)' \\ \beta = \|\widetilde{X}\|_2(\|\widetilde{X}\|_2+|x_j|) \end{cases} \tag{9.82}$$

由此确定出以下矩阵

$$H = I - \beta^{-1}VV' = \begin{bmatrix} I_{j-1} & 0 & 0 \\ 0 & \widetilde{H} & 0 \\ 0 & 0 & I_{n-k} \end{bmatrix} \tag{9.83}$$

其中 $\widetilde{H}$ 为 $k-(j-1)$阶 $H$ 矩阵，容易验证

$$HX = X - V = (x_1,\cdots,x_{j-1},-\sigma_j,0,\cdots,0,x_{k+1},\cdots,x_n)'$$

## §3 雅可比方法

雅可比方法是求实对称矩阵全部特征值及相应特征向量的方法，属于解特征值问题的变换法。在叙述方法前，先注意有关实对称矩阵 $A$ 的以下结论：

① 实对称矩阵 $A$ 的特征值全为实值。

② 若 $A$ 为实对称矩阵，则存在正交矩阵 $V$，可将 $A$ 经相似变换后化为对角矩阵，即

$$V^{-1}AV = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} = D$$

对角矩阵 $D$ 的对角线元素即为 $A$ 的特征值，而 $V$ 的各列即为相应的特征向量。因 $V$ 的各列具

有相互正交的特性，亦即对称矩阵 $\boldsymbol{A}$ 的 $n$ 个特征向量具有两两正交的性质。

③ 若 $\boldsymbol{V}$ 为正交矩阵，则对 $\boldsymbol{A}$ 相似变换后所得的矩阵仍为对称矩阵。

事实上，$\boldsymbol{A}'_1=(\boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{V})'=\boldsymbol{V}'\boldsymbol{A}'(\boldsymbol{V}^{-1})'=\boldsymbol{V}^{-1}\boldsymbol{A}(\boldsymbol{V}')'=\boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{V}=\boldsymbol{A}_1$。

雅可比方法的要点在于：利用正交相似变换后保持 $F$-范数不变的特性，选取一系列旋转相似变换，逐次加大对称矩阵对角元的数值；减少非对角元的数值，从而最终达到将 $\boldsymbol{A}$ 约化为对角阵的目的。具体方法介绍如下。

## 3.1　古典雅可比法

### 3.1.1　方法描述

设 $\boldsymbol{A}$ 为实对称矩阵，记 $\boldsymbol{A}^{(0)}=\boldsymbol{A}$，取 $k=0$，由 $\boldsymbol{A}^{(k)}$ 出发，计算进行如下。

① 选取 $\boldsymbol{A}^{(k)}$ 非对角线的主元素

$$|a_{ij}^{(k)}|=\max_{p\neq q}|a_{pq}^{(k)}| \tag{9.84}$$

由此确定出 $i,j$。

② 按式(9.60)确定 $\varphi_k$，计算出 $\sin\varphi_k$，$\cos\varphi_k$ 值(参看本节 3.1.4)。

③ 对 $\boldsymbol{A}^{(k)}$ 作正交相似变换

$$\boldsymbol{A}^{(k+1)}=(\boldsymbol{V}_{ij}(\varphi_k))'\boldsymbol{A}^{(k)}\boldsymbol{V}_{ij}(\varphi_k) \tag{9.85}$$

按以下公式计算 $\boldsymbol{A}^{(k+1)}$ 的元素：

$$\begin{cases}a_{ii}^{(k+1)}=a_{ii}^{(k)}\cos^2\varphi_k+a_{jj}^{(k)}\sin^2\varphi_k+2a_{ij}^{(k)}\sin\varphi_k\cos\varphi_k\\ a_{jj}^{(k+1)}=a_{ii}^{(k)}\sin^2\varphi_k+a_{jj}^{(k)}\cos^2\varphi_k-2a_{ij}^{(k)}\sin\varphi_k\cos\varphi_k\\ a_{il}^{(k+1)}=a_{li}^{(k+1)}=a_{li}^{(k)}\cos\varphi_k+a_{lj}^{(k)}\sin\varphi_k(i\text{ 行},l\text{ 列元素};l\neq i,j)\\ a_{jl}^{(k+1)}=a_{lj}^{(k+1)}=-a_{li}^{(k)}\sin\varphi_k+a_{lj}^{(k)}\cos\varphi_k(j\text{ 行},l\text{ 列元素};l\neq i,j)\\ a_{ij}^{(k+1)}=a_{ji}^{(k+1)}=0\\ a_{mn}^{(k+1)}=a_{mn}^{(k)}\quad(m,n\neq i,j)\end{cases} \tag{9.86}$$

④ 判断 $|a_{pq}|<\varepsilon$? $(p\neq q)$，若满足，则上述计算过程结束，$\boldsymbol{A}^{(k+1)}$ 的对角线元素就是 $\boldsymbol{A}$ 的特征值的近似值；否则，$k$ 增 1 后转①。

以上简列了由 $\boldsymbol{A}^{(k)}$ 到 $\boldsymbol{A}^{(k+1)}$ 的过程，其中，对于不同的 $k$，选定的下标值 $i$、$j$ 一般也是不同的。虽然可把 $\boldsymbol{A}^{(k)}$ 中的 $a_{ij}^{(k)}$、$a_{ji}^{(k)}$ 置为 0，但在下一步计算中，在该相同位置上可能又变成非 0 元素。因此，雅可比法是一种迭代法。由(9.61)式知，通过一次迭代，$\boldsymbol{A}^{(k+1)}$ 的非对角线元素的平方和减少了 $2(a_{ij}^{(k)})^2$，而对角线元素的平方和却增加了 $2(a_{ij}^{(k)})^2$。由于 $\boldsymbol{A}^{(k)}$ 的对称性，所有的运算都只需在右上三角部分进行。

### 3.1.2　方法的收敛性

**定理 9.1**　设 $\boldsymbol{A}^{(0)}=\boldsymbol{A}$ 为实对称矩阵，对 $\boldsymbol{A}^{(0)}$ 施行一系列旋转变换

$$\boldsymbol{A}^{(k+1)}=(\boldsymbol{V}_{ij}(\varphi_k))'\boldsymbol{A}^{(k)}\boldsymbol{V}_{ij}(\varphi_k)\qquad(k=0,1,2,\cdots)$$

则 $\lim\limits_{k\to\infty}\boldsymbol{A}^{(k)}=\boldsymbol{D}$　　($\boldsymbol{D}$——对角矩阵)　　(9.87)

**证**　记 $\boldsymbol{A}^{(k)}$ 的非对角线元素的平方和为

$$E_k=\sum_{k\neq l}[a_{kl}^{(k)}]^2$$

则

$$E_{k+1}=E_k-2[a_{ij}^{(k)}]^2$$

由(9.84)式知

$$E_k \leqslant n(n-1)[a_{ij}^{(k)}]^2$$

则

$$[a_{ij}^{(k)}]^2 \geqslant \frac{E_k}{n(n-1)}$$

$$E_{k+1} = E_k - 2[a_{ij}^{(k)}]^2 \leqslant E_k - 2\frac{E_k}{n(n-1)} = E_k\left[1 - \frac{2}{n(n-1)}\right]$$

反复应用上式，有

$$E_k \leqslant E_{k-1}\left[1 - \frac{2}{n(n-1)}\right] \leqslant \cdots \leqslant E_0\left[1 - \frac{2}{n(n-1)}\right]^k$$

因为 $1 - \frac{2}{n(n-1)} < 1$，所以 $\lim\limits_{k\to\infty} E_k = 0$。同时还可证明 $\lim\limits_{k\to\infty} \boldsymbol{A}^{(k)}$ 存在，即当 $k\to\infty$ 时，$\boldsymbol{A}^{(k)}$ 以对角阵为其极限，$\lim\limits_{k\to\infty} \boldsymbol{A}^{(k)} = \boldsymbol{D}$。

从以上分析可知，雅可比法产生了一个确定的以对角阵为极限的矩阵序列，而且它们彼此相似，也与原矩阵 $\boldsymbol{A}$ 相似，所以对角阵中对角线上的元素，就是原矩阵的全部特征值。

### 3.1.3 特征向量的计算

设逐次所用的旋转变换矩阵为 $\boldsymbol{V}_{ij}^{(0)}, \boldsymbol{V}_{ij}^{(1)}, \cdots, \boldsymbol{V}_{ij}^{(k)}$。令

$$\boldsymbol{V}_k = \boldsymbol{V}_{ij}^{(0)}\boldsymbol{V}_{ij}^{(1)}\cdots\boldsymbol{V}_{ij}^{(k)} \tag{9.88}$$

因正交矩阵之积仍是正交矩阵，所以 $\boldsymbol{V}_k$ 为正交矩阵。由雅可比法的收敛性知 $\boldsymbol{A}^{(k+1)} = \boldsymbol{V}_k'\boldsymbol{A}\boldsymbol{V}_k \approx \boldsymbol{D}$，因得

$$\boldsymbol{A}\boldsymbol{V}_k \approx \boldsymbol{V}_k\boldsymbol{D} \tag{9.89}$$

设 $\boldsymbol{V}_k = (\boldsymbol{v}_1^{(k)}, \boldsymbol{v}_2^{(k)}, \cdots, \boldsymbol{v}_n^{(k)})$，其中 $\boldsymbol{v}_j^{(k)}$ $(j=1,2,\cdots,n)$ 为 $\boldsymbol{V}_k$ 的列向量。将 $\boldsymbol{V}_k$ 代入式(9.89)得

$$(\boldsymbol{A}\boldsymbol{v}_1^{(k)}, \boldsymbol{A}\boldsymbol{v}_2^{(k)}, \cdots, \boldsymbol{A}\boldsymbol{v}_n^{(k)}) \approx (\boldsymbol{v}_1^{(k)}, \boldsymbol{v}_2^{(k)}, \cdots, \boldsymbol{v}_n^{(k)})\begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix} = (\lambda_1\boldsymbol{v}_1^{(k)}, \lambda_2\boldsymbol{v}_2^{(k)}, \cdots, \lambda_n\boldsymbol{v}_n^{(k)})$$

可见矩阵 $\boldsymbol{V}_k$ 的列向量便是与 $\boldsymbol{D}$ 中主对角线元素相应的特征向量。

由式(9.88)可获得以下递推公式

$$\boldsymbol{V}_k = (\boldsymbol{V}_{ij}^{(0)}\boldsymbol{V}_{ij}^{(1)}\cdots\boldsymbol{V}_{ij}^{(k-1)})\boldsymbol{V}_{ij}^{(k)} = \boldsymbol{V}_{k-1}\boldsymbol{V}_{ij}^{(k)} \tag{9.90}$$

记 $\boldsymbol{V}_k = [v_{ij}^{(k)}]$，则 $\boldsymbol{V}_k$ 的元素计算公式如下

$$\begin{cases} v_{pl}^{(k)} = v_{pl}^{(k-1)} & (l \neq i, j) \\ v_{pi}^{(k)} = v_{pi}^{(k-1)}\cos\varphi_k + v_{pj}^{(k-1)}\sin\varphi_k \\ v_{pj}^{(k)} = -v_{pi}^{(k-1)}\sin\varphi_k + v_{pj}^{(k-1)}\cos\varphi_k \end{cases} \tag{9.91}$$

### 3.1.4 $\sin\varphi_k$、$\cos\varphi_k$ 的计算

在式(9.86)与式(9.91)中，需要使用 $\sin\varphi_k$ 与 $\cos\varphi_k$ 的数值，它们的计算公式可由下式

$$\tan 2\varphi_k = \frac{2a_{ij}^{(k)}}{a_{ii}^{(k)} - a_{jj}^{(k)}} \tag{9.92}$$

导出。常将 $2\varphi_k$ 限制在以下范围

$$-\frac{\pi}{2} \leqslant 2\varphi_k \leqslant \frac{\pi}{2}, \quad 或 \quad -\frac{\pi}{4} \leqslant \varphi_k \leqslant \frac{\pi}{4} \tag{9.93}$$

因此，若 $a_{ii}^{(k)}-a_{jj}^{(k)}=0$ 时，则取

$$\varphi_k=\begin{cases}-\dfrac{\pi}{4}, & a_{ij}^{(k)}<0\\[2mm] \dfrac{\pi}{4}, & a_{ij}^{(k)}>0\end{cases} \tag{9.94}$$

令　　$y=|a_{ii}^{(k)}-a_{jj}^{(k)}|$，　$x=\mathrm{sgn}(a_{ii}^{(k)}-a_{jj}^{(k)})\cdot 2a_{ij}^{(k)}$

则有

$$\begin{cases}\tan 2\varphi_k=\dfrac{x}{y}\\[2mm] \cos 2\varphi_k=\dfrac{1}{\sqrt{1+\tan^2 2\varphi_k}}=\dfrac{1}{\sqrt{1+\left(\dfrac{x}{y}\right)^2}}=\dfrac{y}{\sqrt{x^2+y^2}}\\[2mm] \sin 2\varphi_k=\tan 2\varphi_k\cdot\cos 2\varphi_k=\dfrac{x}{y}\cdot\dfrac{y}{\sqrt{x^2+y^2}}=\dfrac{x}{\sqrt{x^2+y^2}}\end{cases} \tag{9.95}$$

$$\begin{cases}\cos\varphi_k=\sqrt{\dfrac{1}{2}(1+\cos 2\varphi_k)}=\sqrt{\dfrac{1}{2}\left(1+\dfrac{y}{\sqrt{x^2+y^2}}\right)}\\[2mm] \sin\varphi_k=\dfrac{\sin 2\varphi_k}{2\cos\varphi_k}=\dfrac{x}{2\cos\varphi_k\sqrt{x^2+y^2}}\end{cases} \tag{9.96}$$

**例 9.5**　用雅可比法计算下列对称矩阵

$$\boldsymbol{A}=\begin{bmatrix}4 & 2 & 2\\ 2 & 5 & 1\\ 2 & 1 & 6\end{bmatrix}$$

的全部特征值。

**解**　记 $\boldsymbol{A}^{(0)}=\boldsymbol{A}$，选 $a_{12}^{(0)}=2$ 为 $\boldsymbol{A}^{(0)}$ 的非对角线元素的主元素，这时 $i=1$、$j=2$。按式(9.96)计算

$$y=|a_{11}^{(0)}-a_{22}^{(0)}|=|4-5|=1$$

$$x=\mathrm{sgn}(a_{11}^{(0)}-a_{22}^{(0)})\cdot 2a_{12}^{(0)}=(-1)\times 2\times 2=-4$$

$$\cos\varphi_0=\sqrt{\frac{1}{2}\left(1+\frac{1}{\sqrt{(-4)^2+1^2)}}\right)}=0.788\ 206$$

$$\sin\varphi_0=\frac{-4}{2\times 0.788\ 206\sqrt{(-4)^2+1^2}}=-0.615\ 412$$

按式(9.86)计算 $\boldsymbol{A}^{(1)}$ 的元素得

$$a_{11}^{(1)}=a_{11}^{(0)}\cos^2\varphi_0+a_{22}^{(0)}\sin^2\varphi_0+2a_{12}^{(0)}\sin\varphi_0\cos\varphi_0=2.438\ 448$$

$$a_{22}^{(1)}=a_{11}^{(0)}\sin^2\varphi_0+a_{22}^{(0)}\cos^2\varphi_0-2a_{12}^{(0)}\sin\varphi_0\cos\varphi_0=6.561\ 552$$

$$a_{33}^{(1)}=a_{33}^{(0)}=6$$

$$a_{12}^{(1)}=a_{21}^{(1)}=0$$

$$a_{13}^{(1)}=a_{31}^{(1)}=a_{31}^{(0)}\cos\varphi_0+a_{32}^{(0)}\sin\varphi_0=0.961$$

$$a_{23}^{(1)}=a_{32}^{(1)}=-a_{31}^{(0)}\sin\varphi_0+a_{32}^{(0)}\cos\varphi_0=2.020\ 190\ 3$$

所以 $\boldsymbol{A}^{(1)}=\begin{bmatrix}2.438\ 448 & 0 & 0.961\\ 0 & 6.561\ 552 & 2.020\ 190\\ 0.961 & 2.020\ 190 & 6\end{bmatrix}$

继续选 $a_{23}^{(1)}=2.020\ 190$ 为 $\boldsymbol{A}^{(1)}$ 非对角线元素的主元素，即 $i=2,j=3$，仿上计算可得

$$\boldsymbol{A}^{(2)}=\begin{bmatrix}2.438\ 448 & 0.631\ 026 & 0.724\ 794\\ 0.631\ 026 & 8.320\ 386 & 0\\ 0.724\ 794 & 0 & 4.241\ 166\end{bmatrix}$$

$$\boldsymbol{A}^{(3)}=\begin{bmatrix}2.183\ 185 & 0.595\ 192 & 0\\ 0.595\ 192 & 8.320\ 386 & 0.209\ 614\\ 0 & 0.209\ 614 & 4.496\ 424\end{bmatrix}$$

$$\boldsymbol{A}^{(4)}=\begin{bmatrix}2.125\ 995 & 0 & -0.020\ 048\\ 0 & 8.377\ 576 & 0.208\ 653\\ -0.020\ 048 & 0.208\ 653 & 4.496\ 424\end{bmatrix}$$

$$\boldsymbol{A}^{(5)}=\begin{bmatrix}2.125\ 995 & -0.001\ 073 & -0.020\ 019\\ -0.001\ 073 & 8.388\ 761 & 0\\ -0.020\ 019 & 0 & 4.485\ 239\end{bmatrix}$$

$$\boldsymbol{A}^{(6)}=\begin{bmatrix}2.125\ 825 & -0.001\ 072 & 0\\ -0.001\ 072 & 8.388\ 761 & 0.000\ 009\\ 0 & 0.000\ 009 & 4.485\ 401\end{bmatrix}$$

$$\boldsymbol{A}^{(7)}=\begin{bmatrix}2.125\ 825 & 0 & 0\\ 0 & 8.388\ 761 & 0.000\ 009\\ 0 & 0.000\ 009 & 4.485\ 401\end{bmatrix}$$

故 $\boldsymbol{A}$ 的特征值为

$$\lambda_1\approx 2.125\ 825,\quad \lambda_2\approx 8.388\ 761,\quad \lambda_3\approx 4.485\ 401$$

## 3.2 实用雅可比法

上面给出的雅可比法，由于每一次旋转变换都需要寻找非对角线元素的主元素，比较费时。因此在实际使用时，提出了不少修改方案，下面介绍两种可供选择的策略。

### 3.2.1 循环雅可比法

该法不必寻找主元素，而是按照矩阵元素的自然排列次序，由于 $\boldsymbol{A}$ 对称，只要依次地把非对角线元素例如按行的次序 $a_{12},\cdots,a_{1n};a_{23},\cdots,a_{2n};\cdots;a_{(n-1)n}$ 逐次化为零。每作一次化零工作称为一次扫描。一次扫描后，原已化为零的元素可能再次变为非零，需要再次扫描，直至全部非对角线元素达到充分小为止。这个方法有一个明显的缺点，就是把已经很小的元素还要化为零，这实际上并无此必要。因此有下面的改进方法。

### 3.2.2 雅可比过关法

这种方法是先确定一个阈值 $\varepsilon_1$，就 $\varepsilon_1$ 进行扫描，碰到绝对值小于 $\varepsilon_1$ 的非对角线元素就跳过去；否则作化零工作。如此反复循环进行，直到所有非对角线元素的绝对值都小于 $\varepsilon_1$ 为止。然后取 $\varepsilon_2(<\varepsilon_1)$ 作为新的阈值重复上述过程，最后直至全部非对角线元素的绝对值均小于最小的阈值 $\varepsilon_m(<\varepsilon_{m-1}<\cdots<\varepsilon_2<\varepsilon_1)$ 为止。这种方法亦称为限值雅可比法或阈值雅可比法。它已证明是收敛的。

雅可比方法具有收敛快、算法稳定的优点，而且求得的特征向量有很好的正交性；缺点是运算量大。它是适用于求解中小型实对称矩阵全部特征值和特征向量的较好方法。

# §4　QR 方法

QR 方法是近代发展起来的求任意方阵全部特征值及其特征向量的最有效的方法，和雅可比方法一样，它也要对原矩阵进行一系列的正交相似变换，只是 QR 方法是先进行 QR 分解，然后通过逆序相乘来实现正交相似变换的。至于 QR 分解可以用旋转变换或豪斯荷尔德变换来实现。

## 4.1　矩阵的 QR 分解

**定理 9.2**　设 $A\in\mathbf{R}^{n\times n}$，则存在正交矩阵 $P$，使 $PA=R$，其中 $R$ 为右上三角矩阵。

**证**　我们用构造性的方法进行证明。对矩阵 $A$，可按(9.54)式确定 $V_{12}$，则 $V_{12}A$ 可将 $A$ 的(2,1)位置上的元素化为 0。同法确定 $V_{13},V_{14},\cdots,V_{1n}$，继续逐次左乘，则逐次变换后的矩阵在(3,1)、(4,1)、…、$(n,1)$位置上的元素为 0。记

$$P_1=V_{1n}V_{1(n-1)}\cdots V_{12} \tag{9.97}$$

则 $P_1A$ 的第一列对角元以下的元素一定为 0。

同理可找到

$$P_2=V_{2n}V_{2(n-1)}\cdots V_{23} \tag{9.98}$$

使 $P_2P_1A$ 第二列对角元以下的元素为 0，而第一列对角元以下元素与 $P_1A$ 一样是 0。逐步计算可得

$$P_{n-1}P_{n-2}\cdots P_2P_1A=R \tag{9.99}$$

式中，$R$ 为右上三角阵，$P=P_{n-1}P_{n-2}\cdots P_2P_1$ 为正交矩阵。(证毕)

我们也可以用 $H$ 矩阵来构造正交矩阵 $P$。记 $A^{(0)}=A$，它的第一列记为 $a_1^{(0)}$，按式(9.75)～式(9.78)，可找到 $H_1\in\mathbf{R}^{n\times n}$，使

$$H_1a_1^{(0)}=k_1e_1 \tag{9.100}$$

令 $A^{(1)}=H_1A^{(0)}$，其第一列除对角元外均为 0。一般地设

$$A^{(j-1)}=\begin{bmatrix}D^{(j-1)} & B^{(j-1)}\\ 0 & \widetilde{A}^{(j-1)}\end{bmatrix} \tag{9.101}$$

式中，$D^{(j-1)}$ 为 $j-1$ 阶方阵，其对角线以下元素已是 0。$\widetilde{A}^{(j-1)}$ 为 $n-(j-1)$ 阶方阵，其第一列设为 $a_1^{(j-1)}$，可选择 $n-(j-1)$ 阶的 $\widetilde{H}_j$ 矩阵，使

$$\widetilde{H}_ja_1^{(j-1)}=k_j(1,0,\cdots,0)'\in\mathbf{R}^{n-(j-1)} \tag{9.102}$$

根据 $\widetilde{H}_j$ 构造 $H_j$ 矩阵

$$H_j=\begin{bmatrix}I_{j-1} & 0\\ 0 & \widetilde{H}_j\end{bmatrix} \tag{9.103}$$

就有 $A^{(j)}=H_jA^{(j-1)}=\begin{bmatrix}D^{(j)} & B^{(j)}\\ 0 & \widetilde{A}^{(j)}\end{bmatrix}$　(9.104)

它和 $A^{(j-1)}$ 有类似的形式，只是 $D^{(j)}$ 为 $j$ 阶方阵，其对角线以下元素是 0。如此经$(n-1)$步运算后得到

$$H_{n-1}H_{n-2}\cdots H_1A=A^{(n-1)}=R \tag{9.105}$$

式中，$R=A^{(n-1)}$ 为右上三角矩阵，$H=H_{n-1}\cdots H_1$ 为正交矩阵。(证毕)

**定理 9.3($QR$ 分解定理)** 设 $A\in\mathbf{R}^{n\times n}$ 为非奇异矩阵,则存在正交矩阵 $Q$ 与右上三角矩阵 $R$,使 $A$ 有如下分解

$$A = QR$$

且当 $R$ 的对角元符号取定时,分解是唯一的。

**证** 由式(9.99)或式(9.105)知

$$A = P^{-1}R = P'R \quad 或 \quad A = H^{-1}R = H'R \tag{9.106}$$

只要令正交矩阵 $Q=P'$ 或 $Q=H'$,就有 $A=QR$。

为说明分解的唯一性,设有两种分解

$$A = Q_1R_1 = Q_2R_2 \tag{9.107}$$

因 $A$ 非奇异,则 $R_1$、$R_2$ 亦非奇异。由此可得

$$Q_2^{-1}Q_1 = R_2R_1^{-1}$$

或

$$Q_2'Q_1 = R_2R_1^{-1} \tag{9.108}$$

因为右上三角阵的逆阵及两个右上三角阵的乘积均是右上三角阵,故 $R_2R_1^{-1}$ 是右上三角阵。而 $Q_2'$ 与 $Q_1$ 为正交矩阵,其积 $Q_2'Q_1$ 仍是正交矩阵。由式(9.108)知,$R_2R_1^{-1}$ 是右上三角正交矩阵,据正交矩阵性质得

$$(R_2R_1^{-1})' = (R_2R_1^{-1})^{-1} \tag{9.109}$$

这个式子左边是左下三角阵,而右边为右上三角阵的逆阵,它仍是右上三角阵。因此若要式(9.109)的等式成立,$R_2R_1^{-1}$ 只能是对角阵,设

$$R_2R_1^{-1} = D = \mathrm{diag}(d_1, d_2, \cdots, d_n) \tag{9.110}$$

$D$ 为对角正交矩阵,其对角元只能为+1 或−1。综上可得

$$Q_1 = Q_2D, \quad DR_1 = R_2 \tag{9.111}$$

从上式可见,当 $D$ 中的对角元有负号出现时,则 $Q_2$ 与 $Q_1$ 的相应列及 $R_2$ 与 $R_1$ 的相应行可以不同于一个负号,因此 $QR$ 分解不是唯一的。但当 $R_1$ 与 $R_2$ 的对角元取定时,则由

$$DR_1 = R_2$$

可得

$$d_ir_{ii}^{(1)} = r_{ii}^{(2)} \qquad (i=1,2,\cdots,n) \tag{9.112}$$

式中,$r_{ii}^{(1)}, r_{ii}^{(2)}\ (i=1,2,\cdots,n)$ 分别是 $R_1$、$R_2$ 的对角元。由定理条件 $r_{ii}^{(1)}$ 与 $r_{ii}^{(2)}$ 同号及式(9.112)推得 $d_i>0$,又因 $D$ 为对角正交矩阵,所以 $d_i=1(i=1,2,\cdots,n)$,由此知 $D=I$,据式(9.111)就有

$$R_1 = R_2, \quad Q_1 = Q_2 \tag{9.113}$$

成立,定理得证。

定理中所指的唯一性是针对 $R$ 的对角元符号取定的情况而言的。若约定 $R$ 的对角元均取正值,则在这种约定下的 $QR$ 分解同样是唯一的。一般按定理 9.2 中所述的变换方法作出的 $QR$ 分解,$R$ 的对角元不一定全是正的。若令

$$A = (Q\tilde{D}^{-1})(\tilde{D}R) = \tilde{Q}\tilde{R} \tag{9.114}$$

其中

$$\begin{cases} \tilde{D} = \mathrm{diag}\left(\dfrac{r_{11}}{|r_{11}|}, \dfrac{r_{22}}{|r_{22}|}, \cdots, \dfrac{r_{nn}}{|r_{nn}|}\right) \\ r_{ii}——R \text{ 的对角元} \end{cases} \tag{9.115}$$

则 $\tilde{\boldsymbol{R}}$ 便是对角元均为正的右上三角阵了。

**例 9.6**　用 $\boldsymbol{H}$ 变换矩阵作 $\boldsymbol{A}$ 的 $QR$ 分解。

$$\boldsymbol{A}=\begin{bmatrix}2 & -2 & 3\\ 1 & 1 & 1\\ 1 & 3 & -1\end{bmatrix}$$

**解**　按式(9.75)～式(9.78)找 $\boldsymbol{H}_1\in\mathbf{R}^{3\times3}$,则有

$$\boldsymbol{H}_1=\begin{bmatrix}-0.816\,497 & -0.408\,248 & -0.408\,248\\ -0.408\,248 & 0.908\,248 & -0.091\,751\,7\\ -0.408\,248 & -0.091\,751\,7 & 0.908\,248\end{bmatrix}$$

$$\boldsymbol{H}_1\boldsymbol{A}=\begin{bmatrix}-2.449\,49 & 0 & -2.449\,49\\ 0 & 1.449\,49 & -0.224\,745\\ 0 & 3.449\,49 & -2.224\,74\end{bmatrix}$$

再找 $\tilde{\boldsymbol{H}}_2\in\mathbf{R}^{2\times2}$得

$$\tilde{\boldsymbol{H}}_2=\begin{bmatrix}-0.387\,392 & -0.921\,915\\ -0.921\,915 & 0.387\,392\end{bmatrix}$$

$$\boldsymbol{H}_2=\begin{bmatrix}1 & 0\\ 0 & \tilde{\boldsymbol{H}}_2\end{bmatrix}=\begin{bmatrix}1 & 0 & 0\\ 0 & -0.387\,392 & -0.921\,915\\ 0 & -0.921\,915 & 0.387\,392\end{bmatrix}$$

$$\boldsymbol{H}_2(\boldsymbol{H}_1\boldsymbol{A})=\begin{bmatrix}-2.449\,49 & 0 & -2.449\,49\\ 0 & -3.741\,66 & 2.138\,09\\ 0 & 0 & -0.654\,654\end{bmatrix}=\boldsymbol{R}$$

这是一个右上三角阵,但对角元皆为负数,可令 $\tilde{\boldsymbol{D}}=-\boldsymbol{I}$,则$\tilde{\boldsymbol{R}}=-\boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{A}$ 是对角元为正的右上三角阵,而

$$\tilde{\boldsymbol{Q}}=(\boldsymbol{H}_2\boldsymbol{H}_1)'(-\boldsymbol{I})=-(\boldsymbol{H}_2\boldsymbol{H}_1)'=\begin{bmatrix}0.816\,497 & -0.534\,522 & -0.218\,218\\ 0.408\,248 & 0.267\,261 & 0.872\,872\\ 0.408\,248 & 0.801\,783 & -0.436\,436\end{bmatrix}$$

## 4.2　*QR* 算法

### 4.2.1　基本 *QR* 算法

在分解定理 9.3 的基础上,可设计如下 $QR$ 算法。

令 $\boldsymbol{A}_1=\boldsymbol{A}$,按定理 9.2 的方法进行正交分解,分解为正交矩阵 $\boldsymbol{Q}_1$ 和右上三角阵 $\boldsymbol{R}_1$ 的乘积

$$\boldsymbol{A}_1=\boldsymbol{Q}_1\boldsymbol{R}_1$$

然后将得到的因式矩阵 $\boldsymbol{Q}_1$ 和 $\boldsymbol{R}_1$ 逆序相乘,得

$$\boldsymbol{A}_2=\boldsymbol{R}_1\boldsymbol{Q}_1$$

这样就完成了 $QR$ 算法的一步。以下再以 $\boldsymbol{A}_2$ 代替 $\boldsymbol{A}_1$,重复上述步骤即可得到 $\boldsymbol{A}_3$。如此类推,便得 $QR$ 算法的计算公式为

$$\begin{cases}\boldsymbol{A}_k=\boldsymbol{Q}_k\boldsymbol{R}_k\\ \boldsymbol{A}_{k+1}=\boldsymbol{R}_k\boldsymbol{Q}_k=\boldsymbol{Q}_{k+1}\boldsymbol{R}_{k+1}\quad(k=1,2,\cdots)\end{cases}\tag{9.116}$$

因 $R_k=Q_k^{-1}A_k$，所以

$$A_{k+1}=Q_k^{-1}A_kQ_k \quad (k=1,2,\cdots) \tag{9.117}$$

可见由 $QR$ 算法产生的矩阵序列 $\{A_k\}$ 中的每一个矩阵 $A_k$ 都与矩阵 $A$ 相似，因此它们有完全相同的特征值。

**4.2.2 QR 方法的收敛性**

传统的矩阵序列 $\{A_k\}$ 的收敛性，是每个元素的数列 $\{a_{ij}^{(k)}\}$ 都收敛。在讨论 $QR$ 方法的收敛性时，为了求得特征值，只要求序列 $\{A_k\}$ 收敛于一种简单形式的矩阵，例如三角形（或分块三角形）矩阵，而其对角线元（或块）有确定的极限即可。因此，就求取特征值而言，可以这样约定，只要 $\{A_k\}$ 收敛于三角形（或分块三角形）矩阵，其对角线元（或子块）有确定的极限，无论其对角线（或子块）外的元素是否有确定极限，都叫做方法是收敛的，这种收敛性亦叫"基本收敛"或"本质收敛"。$QR$ 方法无论在理论上或应用上都是相当复杂的，它的收敛性有各种情况，下面给出特征值按模不相等情况的收敛性结果。

**定理 9.4** 对 $A\in \mathbf{R}^{n\times n}$，设其特征值满足

$$|\lambda_1|>|\lambda_2|>\cdots>|\lambda_n|>0 \tag{9.118}$$

$\lambda_i$ 对应的特征向量为 $X_i(i=1,2,\cdots,n)$。以 $X_i$ 为列的方阵记为 $X=(X_1,X_2,\cdots,X_n)$，设 $X^{-1}$ 可分解为

$$X^{-1}=LU \tag{9.119}$$

式中，$L$ 为单位左下三角阵，$U$ 为右上三角阵。则 $QR$ 算法产生的序列 $\{A_k\}$ 基本收敛于依次以 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 为对角元的右上三角阵，其对角元极限为

$$\lim_{k\to\infty} a_{ii}^{(k)}=\lambda_i \quad (i=1,2,\cdots,n) \tag{9.120}$$

若 $A$ 为对称矩阵，则 $\{A_k\}$ 收敛到对角阵。

这个定理的证明十分复杂，而且条件(9.119)通常是难于验证的，这里不再讨论。对于特征值不满足(9.118)条件的矩阵，序列 $\{A_k\}$ 可能不收敛到三角阵。这里列出与 $QR$ 方法收敛性有关的一些结论。

① 对于任意方阵 $A$，$QR$ 方法所产生的序列 $\{A_k\}$ 将基本收敛于分块右上三角形矩阵，其对角线上每一子块有等模的特征值。

② 如果矩阵 $A$ 等模的各特征值中，只有实的重特征值或复共轭的重特征值对，则上述的对角线子块将收敛于右上三角形或 $2\times 2$ 的分块右上三角形矩阵。

③ 若矩阵 $A$ 有若干组等模但互不相等的特征值，一般情况下，上述的对角线上子块将不收敛于右上三角形或 $2\times 2$ 分块右上三角形矩阵。

**例 9.7** 用 $QR$ 方法求

$$A=\begin{bmatrix} 5 & -2 & -5 & -1 \\ 1 & 0 & -3 & 2 \\ 0 & 2 & 2 & -3 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

的特征值。

**解** 在本例中，$A$ 的特征方程为

$$\lambda^4-5\lambda^3+7\lambda^2-7\lambda-20=0$$

它可分解成

$$(\lambda+1)(\lambda-4)(\lambda^2-2\lambda+5)=0$$

故特征值为$-1,4,1\pm 2i$

用 $QR$ 方法，得到

$$A_{12}=\left[\begin{array}{c:cc:c} 4.0000 & 5.0484 & -3.6564 & * \\ \hdashline 0 & 1.8789 & -3.5910 & * \\ 0 & 1.3290 & 0.1211 & * \\ \hdashline 0 & 0 & 0 & -1.0000 \end{array}\right]$$

略去 * 数值，由此得到一个特征值为 4，另一个为$-1$，还有两个要解下列方程

$$\begin{vmatrix} 1.8789-\lambda & -3.5910 \\ 1.3290 & 0.1211-\lambda \end{vmatrix}=0$$

得 $1\pm 2i$。

在求得 $A$ 的特征值的基础上，然后用反幂法求取其相应的特征向量。

在 $QR$ 方法中，每一步都要进行一次 $Q$、$R$ 分解，再作一次矩阵乘法。因此对一般的实矩阵，其计算量很大。为了减少计算量，通常先作相似变换将原矩阵 $A$ 化为一个拟上三角阵（次对角线以下元素全为 0），即上海森伯格（Hessen berg）阵，它的形式为

$$\begin{bmatrix} \times & \times & \times & \cdots & \times \\ \times & \times & \times & \cdots & \times \\ & \times & \times & \cdots & \times \\ & 0 & \ddots & \ddots & \vdots \\ & & & \times & \times \end{bmatrix} \tag{9.121}$$

然后对它应用 $QR$ 方法，下面介绍变换方法。

### 4.2.3　正交相似变换化矩阵为上海森伯格阵

(1) 旋转相似变换法

如果在(9.56)式中，取 $\psi=\varphi$，则得到 $A$ 的相似变换矩阵

$$A^{(1)}=V_{ij}'(\varphi)AV_{ij}(\varphi) \tag{9.122}$$

由式(9.57)第 3 式知

$$a_{jl}^{(1)}=-a_{il}\sin\varphi+a_{jl}\cos\varphi \tag{9.123}$$

若取 $l=i-1$ 及

$$\begin{cases} \sin\varphi=\dfrac{a_{j(i-1)}}{\sqrt{a_{i(i-1)}^2+a_{j(i-1)}^2}},\quad \cos\varphi=\dfrac{a_{i\,i-1}}{\sqrt{a_{i(i-1)}^2+a_{j(i-1)}^2}} & \left(\sqrt{a_{i(i-1)}^2+a_{j(i-1)}^2}\neq 0\right) \\ \sin\varphi=0,\ \cos\varphi=1 \quad \left(\sqrt{a_{i(i-1)}^2+a_{j(i-1)}^2}=0\right) \end{cases} \tag{9.124}$$

则可使 $a_{j(i-1)}^{(1)}=0$。

按照上述公式，依次确定 $V_{23},V_{24},\cdots,V_{2n}$ 对 $A$ 作相似变换，分别消去(3,1)，(4,1)，…，($n$,1)位置上的元素；继续再用 $V_{34},V_{35},\cdots,V_{3n}$ 作相似变换分别消去(4,2)，(5,2)，…，($n$,2)位置上的元素。按照这样的办法和顺序进行正交相似变换，则已经化为 0 的元素在以后的变换过程中不再改变，继续这种过程 $n-2$ 次，就能把 $A$ 化为上海森伯格阵。

(2) 豪斯荷尔德变换法

记 $\boldsymbol{A}^{(0)}=\boldsymbol{A}$，记它的第一列元素组成的向量为 $\boldsymbol{a}_1^{(0)}=(a_{11}^{(0)},a_{21}^{(0)},\cdots,a_{n1}^{(0)})'$，按式(9.79)～式(9.81)确定 $\boldsymbol{H}_1$，使对 $\boldsymbol{A}^{(0)}$ 相似变换后的矩阵 $\boldsymbol{A}^{(1)}=\boldsymbol{H}_1\boldsymbol{A}\boldsymbol{H}_1$ 第一列次对角线以下的元素化为0，这只要在公式中取 $j=2$ 即得。设 $\boldsymbol{A}^{(1)}$ 的第二列元素构成的向量为 $\boldsymbol{a}_2^{(1)}$，按式(9.79)～式(9.81)确定 $\boldsymbol{H}_2$，使对 $\boldsymbol{A}^{(1)}$ 相似变换后的矩阵 $\boldsymbol{A}^{(2)}=\boldsymbol{H}_2\boldsymbol{A}^{(1)}\boldsymbol{H}_2$ 第二列次对角线以下的元素化为0，这只要在公式中取 $j=3$ 即得。如此推作$(n-2)$次便可将矩阵 $\boldsymbol{A}$ 化为上海森伯格阵。

当 $\boldsymbol{A}$ 为对称矩阵时，则由以上两种正交相似变换所得的矩阵仍是对称矩阵，在这种情况下所得的上海森伯格阵被约化为三对角线阵。

基本 $QR$ 算法具有线性收敛速度，运算量也很大，若不采取适当措施，其实用价值是不大的。不过上述两个缺点，目前都有了较好的克服措施，采取这些措施后，$QR$ 方法就成为实际计算中很有效的一种方法。与基本 $QR$ 算法相区别，把采取改进措施以后的 $QR$ 方法称为扩展的 $QR$ 方法。例如，采用化 $\boldsymbol{A}$ 为上海森伯格阵，移位加速，避免复运算的双步 $QR$ 方法等措施，这些内容这里不再具体介绍。

# 习题九

9.1　用幂法计算下列矩阵按模最大的特征值及对应的特征向量，当特征值有两位小数稳定时迭代终止，取 $\boldsymbol{X}_0=(1,1,1)$

$$(1)\ \boldsymbol{A}=\begin{bmatrix}6&2&1\\2&3&1\\1&1&1\end{bmatrix};\qquad (2)\ \boldsymbol{A}=\begin{bmatrix}2&4&6\\3&9&15\\4&16&36\end{bmatrix}$$

9.2　用反幂法求矩阵

$$\boldsymbol{A}=\begin{bmatrix}6&2&1\\2&3&1\\1&1&1\end{bmatrix}$$

最接近于 6 的特征值和特征向量。

9.3　对矩阵

$$\boldsymbol{A}=\begin{bmatrix}4&3&-2\\3&4&-1\\-2&-1&3\end{bmatrix}$$

写出使其第一行第二列的元素约化为 0 的旋转变换矩阵 $\boldsymbol{V}_{12}$，并以其对 $\boldsymbol{A}$ 作正交相似变换。

9.4　用豪斯荷尔德矩阵变换以下向量

(1) $\boldsymbol{a}=(2,2,1)'$；(2) $\boldsymbol{a}=(12,3,4)'$

为$(*,0,0)'$形式。

9.5　用豪斯荷尔德矩阵作正交相似变换，使

$$\boldsymbol{A}=\begin{bmatrix}12&3&4\\3&1&2\\4&2&2\end{bmatrix}$$

约化为三对角线阵。

9.6　用旋转变换矩阵作正交相似变换，使

$$\boldsymbol{A}=\begin{bmatrix}60&12&16&-15\\12&288&309&185\\16&309&312&80\\-15&185&80&-600\end{bmatrix}$$

约化为三对角线阵。

9.7　用雅可比法求矩阵

$$\boldsymbol{A}=\begin{bmatrix}3&-4&3\\-4&6&3\\3&3&1\end{bmatrix}$$

的全部特征值及特征向量。

9.8 把矩阵

$$A = \begin{bmatrix} 12 & -20 & 43 \\ 9 & -15 & -63 \\ 20 & 50 & 35 \end{bmatrix}$$

作 $QR$ 分解。

9.9 用 $QR$ 算法求矩阵

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

的特征值。

# 第十章　快速傅里叶变换

本章简要介绍傅里叶变换的基本概念及性质。在实际问题中，信息分析如频谱分析、数字滤波等都要使用傅里叶变换这一数学工具。傅里叶变换的实质是将时域信号变换为频域信号，从而使信号分析处理大为简化。离散傅里叶(DFT)变换一直是许多工程领域中的基本分析工具，但是DFT在具体数值计算中却遇到了很大困难，因为傅里叶变换是一种正比于$n^2$次的运算过程，计算工作量非常大，即使使用计算速度很快的机器来计算离散傅里叶变换也很耗时，致使有些问题无法通过这一途径来解决。而快速傅里叶变换(FFT)的出现解决了这一矛盾。

快速傅里叶变换(Fast fourier transform)的起源可以追溯到1903年，那时Runge引进了12点和24点的傅里叶变换算法，但Runge的算法没有得到推广。到1942年Daniclson和Lanczos提出了傅里叶变换的最优计算方法，因为当时还没有出现电子计算机，这一有效的方法也没有得到应有的重视。直到1965年由Cooley和Tukey提出了适用于在计算机上求傅里叶变换的快速算法，引起了广泛的重视和应用，也推动了其他各种快速算法的产生，并使有关领域的科学分析面貌为之一新。

下面先介绍一下有限离散傅里叶变换。

## §1　有限离散傅里叶变换

设$f(t)$是定义于$(-\infty,+\infty)$上的函数，满足条件

$$\int_{-\infty}^{+\infty}|f(t)|\,\mathrm{d}t<+\infty \tag{10.1}$$

在上述条件下，对每一实数$\omega$，积分

$$\int_{-\infty}^{+\infty}f(t)\mathrm{e}^{-2\pi\mathrm{i}\omega t}\,\mathrm{d}t \tag{10.2}$$

都存在。这种含参数$\omega$的积分称为傅氏积分。这个积分定义一个函数

$$F(\omega)=\int_{-\infty}^{+\infty}f(t)\mathrm{e}^{-2\pi\mathrm{i}\omega t}\,\mathrm{d}t,\qquad -\infty<\omega<+\infty \tag{10.3}$$

式中，$\mathrm{i}=\sqrt{-1}$，这样就建立了一个把函数$f(t)$变为函数$F(\omega)$的变换，可以证明

$$f(t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}F(\omega)\mathrm{e}^{-2\pi\mathrm{i}\omega t}\,\mathrm{d}w \tag{10.4}$$

称$F(\omega)$为$f(t)$的傅里叶变换。在信息科学中$f(t)$经常表示一个信号随着时间或空间位置的变化，$f$的傅里叶变换$F$代表$f(t)$中的各种频率信号成分。实际应用中经常要将信号从“时域”表示$f(t)$转换为“频域”表示$F(\omega)$并对$F(\omega)$进行适当的处理如滤波等，再将$F(\omega)$转换回$f(t)$，以达到某种技术上的要求，如数据压缩、图像处理、图像识别等。

现在，我们将傅氏积分离散化，通过采样取得$f$的某些离散点集上的值。给定$\Delta t$，取$f(t)$在等距间隔$\Delta t$的$(2n+1)$个观测值

$$f(k\Delta t) \qquad (k=0,\pm1,\pm2,\cdots,\pm n)$$

则当 $n$ 充分大时有

$$F(\omega)\approx\int_{-n\Delta t}^{n\Delta t}f(t)\mathrm{e}^{-2\pi i\omega t}\mathrm{d}t \tag{10.5}$$

上式右端用 $2n$ 个左矩形面积和代替后得

$$F(\omega)\approx\sum_{k=-n}^{n-1}f(k\Delta t)\mathrm{e}^{-2\pi i\omega k\Delta t}\Delta t \tag{10.6}$$

化简为

$$F(\omega)\approx\sum_{k=-n}^{-1}f(k\Delta t)\mathrm{e}^{-2\pi i\omega k\Delta t}\Delta t+\sum_{k=0}^{n-1}f(k\Delta t)\mathrm{e}^{-2\pi i\omega k\Delta t}\Delta t \tag{10.7}$$

对于项 $\sum\limits_{k=-n}^{-1}f(k\Delta t)\mathrm{e}^{-2\pi i\omega k\Delta t}\Delta t$，置 $\omega=\dfrac{j}{n\Delta t}\quad(j=0,1,2,\cdots,n-1)$，就有

$$\begin{aligned}\sum_{k=-n}^{-1}f(k\Delta t)\mathrm{e}^{-2\pi i\omega k\Delta t}\Delta t&=\sum_{k=0}^{n-1}f((k-n)\Delta t)\mathrm{e}^{-2\pi i\omega(k-n)\Delta t}\Delta t\\&=\sum_{k=0}^{n-1}f((k-n)\Delta t)\mathrm{e}^{-2\pi i\frac{j}{n\Delta t}(k-n)\Delta t}\Delta t\\&=\sum_{k=0}^{n-1}f((k-n)\Delta t)\mathrm{e}^{-2\pi i\frac{j}{n\Delta t}k\Delta t}\mathrm{e}^{2\pi i\frac{j}{n\Delta t}n\Delta t}\Delta t\\&=\sum_{k=0}^{n-1}f((k-n)\Delta t)\mathrm{e}^{-2\pi i\frac{j}{n}k}\ \mathrm{e}^{2\pi ij}\Delta t\end{aligned}$$

因为 $\mathrm{e}^{2\pi ij}=\mathrm{e}^{2\pi ji}=\cos2\pi j+i\sin2\pi j=1$，代入上式得

$$\sum_{k=-n}^{-1}f(k\Delta t)\mathrm{e}^{-2\pi i\omega k\Delta t}\Delta t=\sum_{k=0}^{n-1}f((k-n)\Delta t)\mathrm{e}^{-2\pi ijk/n}\Delta t \tag{10.8}$$

于是将 $F(\omega)$ 统一写成

$$\begin{aligned}F\left(\frac{j}{n\Delta t}\right)&\approx\sum_{k=0}^{n-1}f((k-n)\Delta t)\mathrm{e}^{-2\pi ijk/n}\Delta t+\sum_{k=0}^{n-1}f(k\Delta t)\mathrm{e}^{-2\pi ijk/n}\Delta t\\&=\sum_{k=0}^{n-1}A_k\mathrm{e}^{-2\pi ijk/n}\end{aligned} \tag{10.9}$$

这里系数 $A_k=\Big[f(k\Delta t)+f((k-n)\Delta t)\Big]\Delta t$，式(10.9)就是式(10.3)的离散表达式。

给定实的或复的序列 $A_0,A_1,\cdots,A_{n-1}$，称序列 $\{x_j\}$：

$$x_j=\sum_{k=0}^{n-1}A_k\mathrm{e}^{-2\pi ijk/n},\qquad j=0,1,2,\cdots,n-1 \tag{10.10}$$

为序列 $\{A_j\}$ 的离散(有限)傅里叶变换，简称 DFT。

如果知道序列 $\{x_j\}$，也可以通过式(10.10)将 $A_k$ 求出。将式(10.10)两边乘以 $\mathrm{e}^{2\pi ijl/n}$，其中 $l$ 取 0 到 $n-1$ 的正整数，并对 $j$ 从 $0\sim(n-1)$ 求和，有

$$\sum_{j=0}^{n-1}x_j\mathrm{e}^{2\pi ijl/n}=\sum_{j=0}^{n-1}\sum_{k=0}^{n-1}A_k\mathrm{e}^{2\pi ij(l-k)/n}=\sum_{k=0}^{n-1}A_k\sum_{j=0}^{n-1}\mathrm{e}^{2\pi ij(l-k)/n}=\sum_{k=0}^{n-1}A_kg_{k,l}$$

式中，$g_{k,l}=\sum\limits_{j=0}^{n-1}q^j$，$q=\mathrm{e}^{2\pi i(l-k)/n}$，可以看做是一个以 $q=\mathrm{e}^{2\pi i(l-k)/n}$ 为公比、首项为 $q^0=1$ 的等比数列的和，通过等比数列的求和公式，当 $l\neq k$ 时，有

$$g_{k,l}=\sum_{j=0}^{n-1}q^j=\frac{1\cdot(1-q^n)}{1-q}=\frac{1-e^{2\pi i(l-k)}}{1-e^{2\pi i(l-k)/n}}$$

显然上式分母 $1-e^{2\pi i(l-k)/n}\neq 0$，而分子 $1-e^{2\pi i(l-k)}=0$，所以 $g_{k,l}=0$。而当 $l\neq k$ 时，$q=1$，所以 $g_{l,k}=n$。从而有

$$\sum_{j=0}^{n-1}x_j e^{2\pi ijl/n}=\sum_{k=0}^{n-1}A_k g_{k,l}=nA_l$$

即

$$A_l=\frac{1}{n}\sum_{j=0}^{n-1}x_j e^{2\pi ijl/n}\qquad(l=0,\ 1,\ 2,\ \cdots,\ n-1)\tag{10.11}$$

称式(10.11)是式(10.10)的逆变换。它们是使用计算机进行傅里叶分析的主要方法，在数字信号处理、全息技术、光谱和声谱分析等很多领域都有广泛应用。

## §2　快速傅里叶变换

记 $W_n=e^{-2\pi i/n}$，则式(10.10)可以写成

$$x_j=\sum_{k=0}^{n-1}A_k e^{-2\pi ijk/n}=\sum_{k=0}^{n-1}A_k[(e^{-2\pi i/n})^j]^k=\sum_{k=0}^{n-1}A_k(W_n^j)^k$$

$$(j=0,\ 1,\ 2,\cdots,\ n-1)\tag{10.12}$$

式中，$\{A_k\}(k=0,1,\cdots,n-1)$是给定的实的或复的序列。现在我们来分析一下离散傅里叶变换的数值计算问题。如果直接用公式(10.12)计算，需要 $n$ 次复数乘法和 $n$ 次复数加法，若把一次复数乘法和一次复数加法叫做一个操作，那么计算全部 $x_i$ 就需要 $n^2$ 个操作。当 $n$ 很大时，这个计算量是很可观的。因此在计算机出现以前只能对较小的 $n$(如 $n=12$，$n=14$)进行手工计算。即使计算机问世后，$n$ 很大的时候计算的代价也是很大的。这对那些使用离散傅里叶变换处理数据和图像的"实时"计算造成很大的负担，严重地影响了计算速度。直到 20 世纪 60 年代后提出了快速算法之后，速度的问题才得到了解决。

快速算法的思想就是尽量减少乘法次数。用式(10.12)计算 $x_j$ 的 $n$ 个公式中，表面看需要 $n^2$ 个乘法，实际上所有 $e^{-2\pi ijk/n}=W_n^{jk}(j,k=0,1,\cdots,n-1)$中，当 $jk=nN+r$($N$ 为正整数)时，因 $W_n^{jk}=W_n^{nN}\cdot W_n^r=1\cdot W_n^r(r=0,1,2,\cdots,n-1)$，所以只需要对 $n$ 个不同的值 $W_n{}^0,W_n{}^1,\cdots,W_n{}^{n-1}$ 作乘法运算，这样就可以大大减少乘法次数。

设 $n=2^m$，$m$ 是正整数。将式(10.10)中的 $j$ 和 $k$ 用二进制数表示为

$$\begin{cases}j=j_{m-1}2^{m-1}+j_{m-2}2^{m-2}+\cdots+j_1 2^1+j_0=(j_{m-1},j_{m-2},\cdots,j_0)\\ k=k_{m-1}2^{m-1}+k_{m-2}2^{m-2}+\cdots+k_1 2^1+k_0=(k_{m-1},k_{m-2},\cdots,k_0)\end{cases}$$

式中，$j_v$、$k_v$ 取 0 或 1，$v=0,\ 1,\ 2,\ \cdots,\ m-1$。又记

$$x_j=x(j_{m-1},j_{m-2},\cdots,j_1,j_0)$$
$$A_k=A(k_{m-1},k_{m-2},\cdots,k_1,k_0)$$

对于 $\sum_{k=0}^{n-1}$，当 $k$ 值用二进制数$(k_{m-1},k_{m-2},\cdots,k_0)$表示时，可以用以下累加和表示为 $\sum_{k_{m-1}=0}^{1}\sum_{k_{m-2}=0}^{1}\cdots\sum_{k_0=0}^{1}$，例如对于 $m=3$ 的二进制数 $k$ 表示为$(k_2,k_1,k_0)$，那么

$$\sum_{k=0}^{7} k = \sum_{k_2=0}^{1}\sum_{k_1=0}^{1}\sum_{k_0=0}^{1}(k_2,k_1,k_0)$$
$$= \sum_{k_2=0}^{1}\sum_{k_1=0}^{1}[(k_2,k_1,0)+(k_2,k_1,1)]$$
$$= \sum_{k_2=0}^{1}[\sum_{k_1=0}^{1}(k_2,k_1,0)+\sum_{k_1=0}^{1}+(k_2,k_1,1)]$$
$$= \sum_{k_2=0}^{1}[(k_2,0,0)+(k_2,1,0)+(k_2,0,1)+(k_2,1,1)]$$
$$= (0,0,0)+(0,1,0)+(0,0,1)+(0,1,1)+(1,0,0)+(1,1,0)+(1,0,1)+(1,1,1)$$

由此可见,二进制数$(k_{m-1},k_{m-2},\cdots,k_0)$中的数字$k_v(v=0,1,2,\cdots,m-1)$遍取0,1后就可以得到数集$\{0,1,2,\cdots,n-1\}$;同样,二进制数$(j_{m-1},j_{m-2},\cdots,j_0)$中的数字$j_v(v=0,1,2,\cdots,m-1)$遍取0,1后可以得到其数集$\{0,1,2,\cdots,n-1\}$。于是式(10.12)可写成

$$x(j_{m-1},j_{m-2},\cdots,j_1,j_0) = \sum_{k=0}^{n-1}A_kW_n^{jk} = \sum_{k_0=0}^{1}\sum_{k_1=0}^{1}\cdots\sum_{k_{m-1}=0}^{1}A(k_{m-1},k_{m-2},\cdots,k_1,k_0)W_n^p \tag{10.13}$$

式中,$p=jk=(j_{m-1}2^{m-1}+j_{m-2}2^{m-2}+\cdots+j_12^1+j_0)(k_{m-1}2^{m-1}+k_{m-2}2^{m-2}+\cdots+k_12^1+k_0)$

利用$W_n^{a+b}=W_n^aW_n^b$,则$W_n^p$可以写成

$$W_n^p = W_n^{jk} = W_n^{j(k_{m-1}2^{m-1}+k_{m-2}2^{m-2}+\cdots+k_12^1+k_0)} = W_n^{j\cdot k_{m-1}2^{m-1}}\cdot W_n^{j\cdot k_{m-2}2^{m-2}}\cdots W_n^{j\cdot k_0} \tag{10.14}$$

现在我们考虑式(10.14)中的每一项,并利用$(W_n^{2^m})^N=\exp\{-\frac{2\pi i}{2^m}2^mN\}=1$($N$为正整数)进行简化。首先考虑第一项

$$W_n^{j\cdot k_{m-1}2^{m-1}} = W_n^{(j_{m-1}2^{m-1}+j_{m-2}2^{m-2}+\cdots+j_12^1+j_0)\cdot k_{m-1}2^{m-1}}$$
$$=W_n^{2^m(j_{m-1}2^{m-2}\cdot k_{m-1})}\cdot W_n^{2^m(j_{m-2}2^{m-3}\cdot k_{m-1})}\cdots W_n^{2^m(j_1\cdot k_{m-1})}W_n^{2^{m-1}(j_0\cdot k_{m-1})}$$
$$=W_n^{2^{m-1}\cdot j_0\cdot k_{m-1}}$$

类似的,对于式(10.14)中的第二项,有

$$W_n^{j\cdot k_{m-2}2^{m-2}} = W_n^{(j_{m-1}2^{m-1}+j_{m-2}2^{m-2}+\cdots+j_12^1+j_0)\cdot k_{m-2}2^{m-2}}$$
$$= W_n^{2^m(j_{m-1}2^{m-3}\cdot k_{m-2})}\cdot W_n^{2^m(j_{m-2}2^{m-4}\cdot k_{m-2})}\cdots W_n^{2^{m-2}(2j_1+j_0)k_{m-2}}$$
$$= W_n^{2^{m-2}(2j_1+j_0)k_{m-2}}$$

更一般的可以得到

$$W_n^{j\cdot k_{m-i}2^{m-i}} = W_n^{2^{m-i}(2^{i-1}j_{i-1}+\cdots+2j_1+j_0)k_{m-i}} \tag{10.15}$$

应用式(10.15),那么式(10.13)可以简化为

$$x(j_{m-1},j_{m-2},\cdots,j_1,j_0)$$
$$=\sum_{k_0=0}^{1}\{\sum_{k_1=0}^{1}\cdots\{\sum_{k_{m-1}=0}^{1}A(k_{m-1},k_{m-2},\cdots,k_1,k_0)\cdot W_n^{2^{m-1}j_0k_{m-1}}\}\cdot W_n^{2^{m-2}(2j_1+j_0)k_{m-2}}\cdots\}\cdot$$
$$W_n^{2^0(2^{m-1}j_{m-1}+\cdots+2j_1+j_0)k_0} \tag{10.16}$$

显见,式(10.16)的右端是使用$W_n^{\tilde{j}\cdot k_v}$型因子的逐次乘法计算过程,在每一次乘法中当$k_v=0$时,$W_n^{\tilde{j}\cdot k_{vn}}=W_n^0=1$,这时可免去乘因子运算;当$k_v=1$并且$\tilde{j}\neq0$时,这时要做乘因子的

运算。为使计算过程规范化地进行，现引入以下记号

$$\begin{cases} B_0(k_{m-1},k_{m-2},\cdots,k_1,k_0)=A(k_{m-1},k_{m-2},\cdots,k_1,k_0) \\ B_1(j_0,k_{m-2},\cdots,k_1,k_0)=\sum\limits_{k_{m-1}=0}^{1} B_0(k_{m-1},k_{m-2},\cdots,k_1,k_0)W_n^{2^{m-1}j_0k_{m-1}} \\ \qquad =B_0(0,k_{m-2},\cdots,k_1,k_0)+B_0(1,k_{m-2},\cdots,k_1,k_0)W_n^{2^{m-1}j_0} \\ B_2(j_0,j_1,k_{m-3},\cdots,k_1,k_0)=\sum\limits_{k_{m-2}=0}^{1} B_1(j_0,k_{m-2},\cdots,k_1,k_0)W_n^{2^{m-2}(2j_1+j_0)k_{m-2}} \\ \qquad =B_1(j_0,0,k_{m-3},\cdots,k_1,k_0)+B_1(j_0,1,k_{m-3},\cdots,k_1,k_0)W_n^{2^{m-2}(2j_1+j_0)} \\ \cdots \\ B_m(j_0,j_1,\cdots,j_{m-1})=\sum\limits_{k_0=0}^{1} B_{m-1}(j_{m-2},\cdots,j_0,k_0)W_n^{(2^{m-1}j_{m-1}+\cdots+2j_1+j_0)k_0} \\ \qquad =B_{m-1}(j_0,\cdots,j_{m-2},0)+B_{m-1}(j_0,\cdots,j_{m-2},1)W_n^{(2^{m-1}j_{m-1}+\cdots+2j_1+j_0)} \\ x(j_{m-1},j_{m-2},\cdots,j_0)=B_m(j_0,j_1,\cdots,j_{m-1}) \end{cases} \tag{10.17}$$

这样式(10.16)的计算过程就可以逐次递推如下。首先由给定的 $A_k=A(k_{m-1},k_{m-2},\cdots,k_0)$ 出发，对 $k_v(v=0,1,2,\cdots,m-1)$ 遍取 0,1 形成实的或复的序列 $\{B_0(k_{m-1},k_{m-2},\cdots,k_0)\}$。由式(10.17)第 2 式的右端知，$\sum\limits_{k_{m-1}=0}^{1} B_0(k_{m-1},k_{m-2},\cdots,k_0)W_n^{2^{m-1}j_0k_{m-1}}$ 只与 $j_0,k_{m-2},\cdots,k_0$ 的取值有关，所以将它表示为 $B_1(j_0,k_{m-2},\cdots,k_0)$ 对 $j_0,k_{m-2},\cdots,k_0$ 遍取 0,1 后形成实的或复的序列 $\{B_1(j_0,k_{m-2},\cdots,k_0)\}$，余类推，最后即可获得结果序列 $\{x_j\}=\{x(j_{m-1},j_{m-2},\cdots,j_0)\}=\{B_m(j_0,j_1,\cdots,j_{m-1})\}$。综上可见，快速傅里叶变换就是通过遍历 $j_v$、$k_v$ 的各值获取各列的元素以及通过各列相关元素间的逐次递推关系实现求和的计算过程。

下面我们来看一下计算量。

$$B_1=\sum_{k_{m-1}=0}^{1} A(k_{m-1},k_{m-2},\cdots,k_0)W_n^{k_{m-1}2^{m-1}j_0}$$

因序列 $\{A_k\}$ 具有 $n=2^m$ 个，所以计算 $B_1$ 需要 $2^m$ 次乘法运算。同样，由 $B_1$ 出发计算 $B_2$ 也需要 $2^m$ 次乘法运算。如此继续下去，直至计算出 $B_m$ 时，需要的乘法计算量为 $T=m2^m$ 次。因为 $m=\log_2 n$，所以可以写作 $T=n\log_2 n(<n^2)$。可见，当 $n$ 很大时，快速傅里叶变换的计算量比傅里叶变换的计算量少很多。

上述快速傅里叶变换的推演过程或许过于繁琐，但是通过实例示范，可知方法既简单又规范，在计算机上也易于实现。我们以 $n=8,m=3$ 为例，说明一下计算过程。此时

$$j=j_2 2^2+j_1 2^1+j_0=(j_2,j_1,j_0)$$
$$k=k_2 2^2+k_1 2^1+k_0=(k_2,k_1,k_0)$$
$$W_8^{jk}=W_8^{(j_2 2^2+j_1 2^1+j_0)(k_2 2^2+k_1 2^1+k_0)}=W_8^{j_0k_2 2^2}\cdot W_8^{(j_1 2+j_0)k_1\cdot 2}\cdot W_8^{(j_2 2^2+j_1 2+j_0)k_0}$$

于是

$$\begin{aligned} x(j_2,j_1,j_0)&=\sum_{k_0=0}^{1}\sum_{k_1=0}^{1}\sum_{k_2=0}^{1}A(k_2,k_1,k_0)W_n^{jk} \\ &=\sum_{k_0=0}^{1}\sum_{k_1=0}^{1}\sum_{k_2=0}^{1}A(k_2,k_1,k_0)W_8^{j_0k_2 2^2}\cdot W_8^{(j_1 2+j_0)k_1\cdot 2}\cdot W_8^{(j_2 2^2+j_1 2+j_0)k_0} \end{aligned} \tag{10.18}$$

$$= \sum_{k_0=0}^{1}\{\sum_{k_1=0}^{1}[\sum_{k_2=0}^{1}A(k_2,k_1,k_0)W_8^{2^2 j_0 \cdot k_2}]\cdot W_8^{2(j_1 2+j_0)\cdot k_1}\}\cdot W_8^{(j_2 2^2+j_1 2+j_0)\cdot k_0}$$

记 $B_0(k_2,k_1,k_0)=A(k_2,k_1,k_0)$，则

$$B_0(0,0,0)=A(0)$$
$$B_0(0,0,1)=A(1)$$
$$B_0(0,1,0)=A(2)$$
$$B_0(0,1,1)=A(3)$$
$$B_0(1,0,0)=A(4)$$
$$B_0(1,0,1)=A(5)$$
$$B_0(1,1,0)=A(6)$$
$$B_0(1,1,1)=A(7)$$

式(10.18)中的式 $\sum_{k_2=0}^{1}A(k_2,k_1,k_0)W_8^{2^2 j_0 k_2}=\sum_{k_2=0}^{1}B_0(k_2,k_1,k_0)W_8^{(j_0,0,0)k_2}$ 仅与 $j_0,k_1,k_0$ 有关，于是可记为

$$B_1(j_0,k_1,k_0)=\sum_{k_2=0}^{1}A(k_2,k_1,k_0)W_8^{2^2 j_0 k_2}=\sum_{k_2=0}^{1}B_0(k_2,k_1,k_0)W_8^{(j_0,0,0)k_2}$$

进一步有

$$B_1(j_0,k_1,k_0)=B_0(0,k_1,k_0)+B_0(1,k_1,k_0)W_8^{(j_0,0,0)} \tag{10.19}$$

$B_1(j_0,k_1,k_0)$ 中的 $j_0,k_1,k_0$，可以分别取 0、1，这样当 $j_0,k_1,k_0$ 遍取 0、1 之后可以得到 8 个等式如下

$$\begin{cases}B_1(0,0,0)=B_0(0,0,0)+B_0(1,0,0)W_8^0\\ B_1(0,0,1)=B_0(0,0,1)+B_0(1,0,1)W_8^0\\ B_1(0,1,0)=B_0(0,1,0)+B_0(1,1,0)W_8^0\\ B_1(0,1,1)=B_0(0,1,1)+B_0(1,1,1)W_8^0\\ B_1(1,0,0)=B_0(0,0,0)+B_0(1,0,0)W_8^4\\ B_1(1,0,1)=B_0(0,0,1)+B_0(1,0,1)W_8^4\\ B_1(1,1,0)=B_0(0,1,0)+B_0(1,1,0)W_8^4\\ B_1(1,1,1)=B_0(0,1,1)+B_0(1,1,1)W_8^4\end{cases}$$

由此就由给定的序列$\{B_0(k_2,k_1,k_0)\}$，计算出了序列$\{B_1(j_0,k_1,k_0)\}$。在式(10.19)基础上，式(10.18)可以化成：

$$x(j_2,j_1,j_0)=\sum_{k_0=0}^{1}\{\sum_{k_1=0}^{1}B_1(j_0,k_1,k_0)W_8^{(j_1 2+j_0)k_1\cdot 2}\}\cdot W_8^{(j_2 2^2+j_1 2+j_0)k_0}$$

$$=\sum_{k_0=0}^{1}\{\sum_{k_1=0}^{1}B_1(j_0,k_1,k_0)W_8^{(j_1,j_0,0)k_1}\}\cdot W_8^{(j_2 2^2+j_1 2+j_0)k_0} \tag{10.20}$$

同样，式 $\sum_{k_1=0}^{1}B_1(j_0,k_1,k_0)W_n^{(j_1,j_0,0)k_1}$ 仅与 $j_0,j_1,k_0$ 有关，于是可记为

$$B_2(j_0,j_1,k_0)=\sum_{k_1=0}^{1}B_1(j_0,k_1,k_0)W_8^{(j_1,j_0,0)k_1}=B_1(j_0,0,k_0)+B_1(j_0,1,k_0)W_8^{(j_1,j_0,0)} \tag{10.21}$$

对于 $B_2(j_0,j_1,k_0)$ 中的 $j_0,j_1,k_0$，遍取 0、1 后同样可以得到以下 8 个等式

$$
\begin{cases}
B_2(0,0,0)=B_1(0,0,0)+B_1(0,1,0)W_8^0 \\
B_2(0,0,1)=B_1(0,0,1)+B_1(0,1,1)W_8^0 \\
B_2(0,1,0)=B_1(0,0,0)+B_1(0,1,0)W_8^4 \\
B_2(0,1,1)=B_1(0,0,1)+B_1(0,1,1)W_8^4 \\
B_2(1,0,0)=B_1(1,0,0)+B_1(1,1,0)W_8^2 \\
B_2(1,0,1)=B_1(1,0,1)+B_1(1,1,1)W_8^2 \\
B_2(1,1,0)=B_1(1,0,0)+B_1(1,1,0)W_8^6 \\
B_2(1,1,1)=B_1(1,0,1)+B_1(1,1,1)W_8^6
\end{cases}
\tag{10.22}
$$

在式(10.22)的基础上，式(10.20)可化为

$$
x(j_2,j_1,j_0)=\sum_{k_0=0}^{1}B_2(j_0,j_1,k_0)\cdot W_8^{(j_2 2^2+j_1 2+j_0)k_0}=\sum_{k_0=0}^{1}B_2(j_0,j_1,k_0)\cdot W_8^{(j_2,j_1,j_0)k_0}
\tag{10.23}
$$

式 $\sum\limits_{k_0=0}^{1}B_2(j_0,j_1,k_0)W_8^{(j_2,j_1,j_0)k_0}$ 仅与 $j_0,j_1,j_2$ 有关，于是可记为

$$
B_3(j_0,j_1,j_2)=\sum_{k_0=0}^{1}B_2(j_0,j_1,k_0)W_8^{(j_2,j_1,j_0)k_0}=B_2(j_0,j_1,0)+B_2(j_0,j_1,1)W_8^{(j_2,j_1,j_0)}
$$

并且有 $B_3(j_0,j_1,j_2)=x(j_2,j_1,j_0)$。

对于 $B_3(j_0,j_1,j_2)$ 中的 $j_0,j_1,j_2$，遍取 0、1 之后同样可以得到 8 个等式如下

$$
\begin{cases}
x(0,0,0)=B_3(0,0,0)=B_2(0,0,0)+B_2(0,0,1)W_8^0 \\
x(1,0,0)=B_3(0,0,1)=B_2(0,0,0)+B_2(0,0,1)W_8^4 \\
x(0,1,0)=B_3(0,1,0)=B_2(0,1,0)+B_2(0,1,1)W_8^2 \\
x(1,1,0)=B_3(0,1,1)=B_2(0,1,0)+B_2(0,1,1)W_8^6 \\
x(0,0,1)=B_3(1,0,0)=B_2(1,0,0)+B_2(1,0,1)W_8^1 \\
x(1,0,1)=B_3(1,0,1)=B_2(1,0,0)+B_2(1,0,1)W_8^5 \\
x(0,1,1)=B_3(1,1,0)=B_2(1,1,0)+B_2(1,1,1)W_8^3 \\
x(1,1,1)=B_3(1,1,1)=B_2(1,1,0)+B_2(1,1,1)W_8^7
\end{cases}
$$

为了能形象地反映以上的计算过程，下面我们把它用流程图表示出来，如图 10.1 所示。流程图可以如下构造：用序号 $j$ 的列作为流程图的左端，然后是列 $B_l(j)(l=0,1,2,3)$，$x$ 列作为流程图的右端。按列的计算过程依次是 $B_0$、$B_1$、$B_2$、$B_3$，每列中的节点用圆圈表示，每个圆圈内部都有一个整数 $z$，表示 $W_n$ 的幂指数，即 $W_n^z$。每一个圆圈都有实线或虚线引入，这表示该节点的值是由实线引来的数与 $W_n^z$ 相乘后再与由虚线引来的数相加，得出该节点的 $B_l(j)$ 值，例如

$$
B_2(1,0,0)=B_1(1,0,0)+B_1(1,1,0)W_8^2
$$

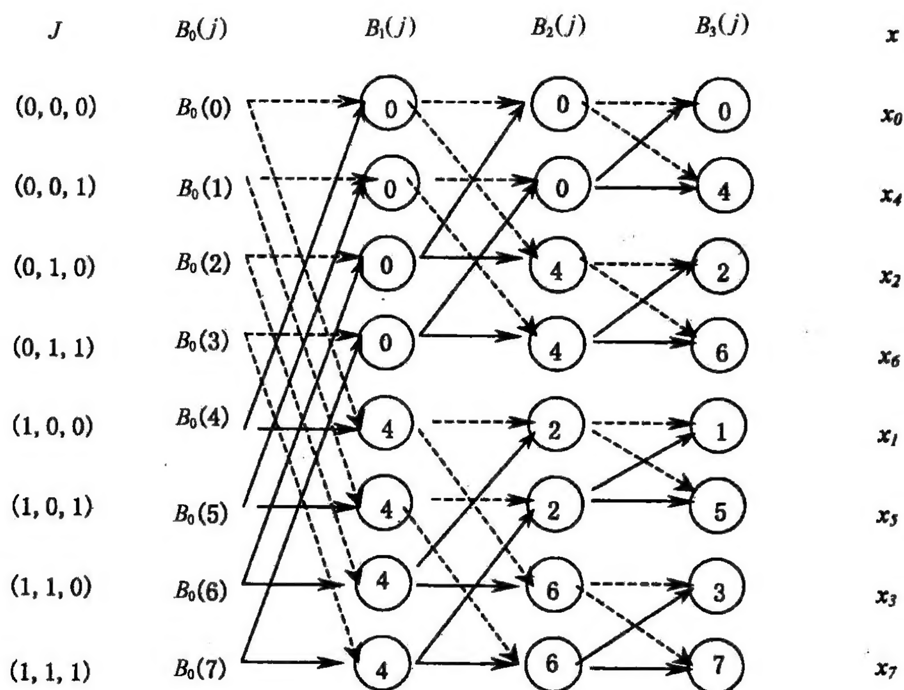

J
$B_0(j)$
$B_1(j)$
$B_2(j)$
$B_3(j)$
x
(0, 0, 0)
(0, 0, 1)
(0, 1, 0)
(0, 1, 1)
(1, 0, 0)
(1, 0, 1)
(1, 1, 0)
(1, 1, 1)
$B_0(0)$
$B_0(1)$
$B_0(2)$
$B_0(3)$
$B_0(4)$
$B_0(5)$
$B_0(6)$
$B_0(7)$
$x_0$
$x_4$
$x_2$
$x_6$
$x_1$
$x_5$
$x_3$
$x_7$

图 10.1

# 习题十

10.1 对 $n=4$ 作快速傅里叶变换流程图。

10.2 当给出的分量 $A_k$ 按照二进制顺序排列时，用快速傅里叶变换方法得到的结果序列 $x_j$ 是乱序的，那么对于 $n=16$ 的情形，请指明乱序的次序是怎样的？

10.3 给定函数 $f(t)=\mathrm{e}^{-|t|}$，求其傅里叶变换。